8-7-2012

Efficient Characterization of Structural Dynamic Responses under Uncertainties: from Order-Reduced Simulation to Data-Driven Emulation

ZEPING XIA
zepingxia@gmail.com

Efficient Characterization of Structural Dynamic Responses under Uncertainties: from Order-Reduced Simulation to Data-Driven Emulation

Zeping Xia

A Thesis

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Master of Science

at the

University of Connecticut

2012

APPROVAL PAGE

Master of Science Thesis

Efficient Characterization of Structural Dynamic Responses under Uncertainties: from Order-Reduced Simulation to Data-Driven Emulation

Presented by

Zeping Xia

Major Advisor _____

Jiong Tang

Associate Advisor _____

Chengyu Cao

Associate Advisor _____

Robert Gao

University of Connecticut

2012

TABLE OF CONTENTS

Abstract

The efficiency of sampling remains as one of the major challenges for uncertainty analysis in structural dynamics. In the context of numerical experiments, this difficulty primarily lies in the fact that every single run of a numerical model, which usually is of large scale in modern engineering practice, would be expensive in terms of CPU times and memories. Nonetheless, understanding and predicting dynamic characteristics of structural systems with uncertainties is important for structural design, assessment and control. In order to develop fast and economical sampling techniques for characterizing structural dynamics, we explored two categories of methods from two opposite direction: the first is first-principle-based order reduced simulation and the second is data based statistical emulation. This thesis particularly focuses on component mode synthesis (CMS) and Gaussian processes regression (GPR).The former is a family of finite-element-based techniques that have drawn significant attention in computational structural dynamics, and the latter is a recently developed statistical approach which has proved to be potentially useful in engineering applications. A general free interface CMS formulation, which has been developed by the author, a concise self-sustained description of GPR and two-level GPR, and a series of application examples will be presented in this thesis.

Chapter 1: Introduction

1.1 Modeling of Structural Dynamics and Uncertainties

Real structural systems are inevitably subject to uncertainties that exist within themselves or in operating environments. Characterization of dynamic behaviors plays an important role in structural design, assessment, and control, especially in cases where uncertainties on structures may lead to significant abnormalities in dynamic responses. Within a structural system, there are various sources of uncertainty, such as material properties, manufacturing tolerance, and in-service degradation. This thesis focuses on discussing this kind of uncertainties that occur within the structure itself.

Even with modern computer capabilities and experimental testing techniques, the characterization of structural dynamic response variation is still a challenging task. In the field of numerical simulation, the dynamics of a structural system are modeled based on underlying physical principles via pure mathematical deductions such as finite element analysis. In this deterministic approach, all causes or the complete information are known for sure about the system and the consequential responses can be worked out accordingly through the deductive logic. Although many efficient numerical algorithms have been available, an accurate dynamic analysis of a complex structure involving a large number of degrees-of-freedom (DOFs) may be still considered expensive in real-world applications. Moreover, a sound analysis of this kind always requires uncertainties to be taken into account because the assumption that the complete information is available never holds in reality.

Uncertainties generally appear to be random phenomena. Some of them such as material property variations may be modeled in terms of probabilities, and others such as geometry variations are difficult to be quantified. There are two basic approaches to introducing uncertainties into structural dynamic analyses. The parametric approach identifies model parameters individually and treats them as random variables within the structural models. The non-parametric approach considers the overall effects of all the uncertainties on the structural dynamical behaviors without specific analyses of individual uncertainties. This thesis only discusses the parametric approach though in practice the implementation of the so-called non-parametric approach essentially requires parametric sampling techniques.

For those parametric uncertainties of a structural system that can be modeled using well-established probability distributions, the characterization of corresponding response variations can be straightforward. A popular approach is to perform uncertainty analyses using deterministic numerical models through Monte Carlo sampling. A large sample of random parameter configurations are drawn from the assumed probability distributions and then passed onto deterministic models for calculating corresponding responses at each sampled parameter configuration. Running brute-force Monte Carlo simulation with large sample size, however, eventually may become computationally impractical as the number of DOFs in a finite element model increases. It should be clear that the parametric approach is based on the assumption that uncertainties can be readily modeled, but the reality is that many uncertainties, such as geometry variation and

boundary condition difference, are difficult to be mathematically described and integrated into numerical models.

1.2 Order Reduction Simulation: CMS

In modern engineering practice, structural dynamic analyses often require large-scale finite element models, especially for complex structures, and running of these models may be computationally expensive. The feasibility of Monte Carlo uncertainty analysis thus becomes questionable when such expensive models are used for sampling. Some researchers have introduced order-reduced models to Monte Carlo uncertainty analysis of structural dynamics (Hinke et al., 2009; Bladh et al., 2002; Lee and Castanier, 2006; Dohnal et al., 2009; Panayirci et al., 2011). Initially order reduction techniques were developed to reduce the computational costs of deterministic finite element analyses in structural dynamics (Tournour et al., 2001; Craig, 2000; Feiner and Griffin, 2002; Moyroud et al., 2002; Mbaye et al., 2010; Bladh et al., 2001; Bah et al., 2003).Among order reduction techniques, component mode synthesis (CMS) is of special interest because it is based on finite element approach, which has become the most popular analyzing tool in industrial applications. The essential idea of CMS is so-called “divide and conquer”. That is to divide a large-scale structure into small components and then build a reduced model for the structure based on analyses on individual components. The reduced model combines retained component mode information together based on connectivity conditions at boundaries and gives birth to approximate solutions which accuracy mostly depend on the number of retained component modes.

Some researchers have introduced CMS models to Monte Carlo uncertainty analysis of structural dynamics (Bladh et al., 2002; Lee and Castanier, 2006; Dohnal et al., 2009). Since order reduction techniques use only the most relevant information extracted from full finite element analyses to reconstruct a small-sized approximate model, obviously there is a trade-off between the computational efficiency and accuracy and the resulting loss of accuracy may lead to sampling errors in Monte Carlo analysis. Another possible issue is that even a reduced model is still too large to be used for Monte Carlo sampling. Furthermore, when physical mechanisms are unknown or too complex to be mathematically modeled confidently, no numerical-model-based analyses can be performed. Only if physical experimental data are available, however, we can carry out analyses using pure data-based methods, such as Gaussian process regression.

1.3 Data-Driven Emulation: GPR

In probability theory and statistics, the Gaussian processes have been explored to establish models for various spatial and temporal problems. The basic idea behind Gaussian processes is to extend the multivariate Gaussian distribution from a finite dimensional space to an infinite dimensional space on which a random continuous function is defined. Over the last decade, the theory of Gaussian process modeling has been gradually formalized by the works on machine learning (Neal, 1998; Mackay, 2003; Rasmussen and Williams, 2006) in such a form that facilitates its applications to modeling of relationship in large high-dimensional data in different backgrounds. Most recently, it has been introduced to the fields of structural dynamics and control

engineering (Hasen et al., 2005; Gregorcic and Lightbody, 2009; Azman and Kocijan, 2009; Mohanty et al., 2009; DiazDelaO and Adhikari, 2010, 2011).In particular, DiazDelaO and Adhikari (2010, 2011) first tried to use Gaussian processes to characterize the structural dynamic responses presumably involving high computational cost.Specifically, they used Gaussian processes as emulators to approximate frequency responses of a simple plate with swept frequencies as inputs of Gaussian process models (DiazDelaO and Adhikari, 2010). Then, they attempted to mitigate difficulties in stochastic finite element methods (DiazDelaO and Adhikari, 2011) by using Gaussian processes to emulate the computationally costly realizations of random fields and implementation of polynomial chaos expansion. While stochastic finite element theory is derived in a rigorous way from probability theory, one major difficulty associated with the numerical incorporation of random processes in finite element analyses lies in the abstractness of domain of the probability space (Ghanem and Spanos, 1991). In essence, all the difficulties encountered in characterization of system response variations/uncertainties are due to the fact that in practice only limited data can be obtained from either physical experiments or numerical realizations so that a plausible uncertainty analysis shall have to resort to inference methods. Recently, with the steady advances in computational capabilities, Bayesian inference has been applied to more and more specific scientific and engineering problems. Bayesian approach gains great popularity in applications primarily because, conceptually, it presents a perfect framework that integrates observations and prior knowledge into a complete formulation which may give rise to plausible predictions of quantities of interest. In the meantime,

within Bayesian framework, Gaussian processes have also been revived and refined as mentioned above accordingly.

1.4 Goals and Structure of This Thesis

In this thesis, the core issue to be addressed is the computational efficiency of Monte Carlo uncertainty analysis for characterizing structural dynamic responses under uncertainties. Two approached are presented from two opposite directions. The first approach starts from well-established physical fundamentals and rigorous mathematical deductions, and the second approach is based on so-called educated guesses about the inner structures of data. Generally speaking, there are three fundamental categories of methodologies to approach a given structural dynamic problem: mechanism-based approach, data-based approach, and approach in between. CMS and GPR, respectively, fall into the first two categories. Perturbation methods are a typical example of the third category where efforts are made to embed statistical concepts in physical formulations.

In Chapter 2, a general formulation of the residual mode for free interface CMS is presented and numerical examples show that CMS is an effective order reducing tool for large complex structures though the truncation error maybe become a deterrent to the use of CMS model for Monte Carlo sampling. The general formulation of residual mode for free-interface component mode synthesis is derived in a concise and straightforward way and, correspondingly, a general formulation of free-interface reduced-order model. Though it is difficult to solve such a nonlinear equation of higher order (>=3), numerical investigations show that in some cases the second order approximation may give results

that are very close to ‘exact’ solutions, e.g., solutions obtained with the full model. In statistics, such discrepancies between results of a CMS model and the full model are termed measurement errors, but there is no general conclusion of their effects on sampling.

In Chapter 3, Gaussian process regression is introduced within Bayesian framework where the parameters of the regression model are determined based on the given dataset. When a system has to be treated as a black box because of the lack of insights on mechanisms behind its behaviors, statistical methods offer quick phenomenological patterns to summarize system behaviors. GPR has great flexibility and learning ability. DiazDelaO and Adhikari (2010, 2010) have applied GPR to the field of stochastic finite element analysis. In this thesis, GPR is employed in a more practical way that it is used to emulate simulations directly. GPR may generate a vast sample using a few data, which can be helpful when sampling costs are high for uncertainty analysis. A spatially periodic structure is chosen for illustration purposes in numerical examples. Such a special structure has dynamic responses that are particularly sensitive to uncertainties. It can be shown that Gaussian process regression may be an efficient Monte Carlo sampling tool if the effect of data errors on the accuracy of inference results is controlled within the acceptable level.

In Chapter 4, In order to improve the inference accuracy while maintain the sampling efficiency, two-level Gaussian process regression is introduced for using data at two different fidelity levels. A common situation in practice is that: given a set of data, for verification and validation purposes we need to acquire a set of reference data, termed

high-fidelity data, which are of high accuracy but, accordingly, usually much more expensive. The given data that need to be verified or validated are termed low-fidelity data. In this thesis, low-fidelity data are provided by CMS models, which are relatively cheap and less accurate as compared to high-fidelity data obtained with full-scale finite element models. In the case where there are many low-fidelity data with errors while a small amount of high-fidelity data are still available, two-level Gaussian process regression is employed for exploiting both the low-fidelity and high-fidelity data simultaneously to emulate high-fidelity sampling with guaranteed accuracy.

It is should be pointed out that, although only numerical examples are used for illustrating the methodologies presented in this thesis, the applications can be readily extended to physical experimental cases.

Chapter 2: Component Mode Synthesis

2.1 A brief overview of CMS

Component mode synthesis (CMS) is a family of order reduction methods that has been under investigation in the last four decades. It has been attractive in structural dynamic analysis while there are still difficulties in obtaining experimental modal data and combining them with numerical analysis data (Misawa, 2010; Butland and Avitabile, 2010; Allen et al., 2010). Essentially, the principle of model reduction is to re-present a full finite element model in a subspace by applying an appropriate transformational basis. CMS provides such a basis by using only a few lowest modes retained in each component/substructure and additional compensating modes for truncated modal information. Based on the different assumptions about boundary conditions at component interfaces, CMS can be classified into two basic categories: fixed interface and free interface methods. The general advantages of the latter over the former include fewer degrees of freedom participating in the synthesized model and, if necessary, easier implementation of experimental testing for modal information. The first step in CMS is to decompose a full structure into substructures and use finite element method to perform modal analysis on each one. Then, a few lowest modes in each substructure are to be retained and some approximation methods are adopted to construct compensating modes for retrieving information from truncated modes. Finally, an order-reduced model for the full structure can be assembled with the retained modes and the compensating modes. A detailed overview of CMS can be found in literature (Craig, 2000; Craig and Kurdila, 2006).

2.2 The general formula of residual flexibility mode

The first order and the second order approximation formula for free interface CMS have been referred to as MacNeal method and Rubin method, respectively (MacNeal, 1971; Rubin, 1975). As a matter of fact, the Rubin method is not the true second order approximation but an approximation which is obtained by iteration using the solutions given by the first order approximation. The true second and third order approximation formulas were proposed by Kang and Kim (Kang and Kim, 2001) and then a concise derivation of the general *n*-th order approximation formula with residual flexibility mode for free interface CMS was proposed recently (Xia and Tang, 2010).

Consider a structure that can be decomposed into substructures. For each substructure, the equation of motion derived based on the finite element method can be written as

$$\mathbf{M}^{(\alpha)}\ddot{\mathbf{x}}^{(\alpha)} + \mathbf{K}^{(\alpha)}\mathbf{x}^{(\alpha)} = \mathbf{f}^{(\alpha)} \tag{1-1}$$

where damping is neglected since our interests lie in eigenvalue problem for modal analysis. **M** and **K** are the *n*-by-*n* mass and stiffness matrices, and **x** and **f** are the physical coordinate vector and the vector of external force applied to the substructure. The superscript α is the index for substructure in question and will be suppressed in following discussions unless otherwise noticed. When Eq (1-1) is partitioned and rearranged in the way that interface DOFs are gathered together, the force vector may take the form

$$\mathbf{f} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f}_J \end{bmatrix} \tag{1-2}$$

where the subscripts J stands for the interface degrees of freedom, and $\mathbf{f}_J$ represents forces at the interfaces due to the interaction between two adjacent substructures. Here, since only free vibration is considered in the eigenvalue problem, the forces external to the substructure have been reduced to the forces at the interfaces only.

As we will see, in the free interface CMS, at the beginning it is assumed that substructures have no connection with each other. Afterwards, in order to represent the physical displacement vector with the kept modes, the interaction at interfaces between adjacent substructures will be taken into account. Finally, the continuity of displacement and force at interfaces will be imposed to form the synthesized global equation. When a substructure is assumed to have no connection with others, no forces is present and Eq (1-1) becomes

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{0} \tag{1-3}$$

Solving the eigenvalue problem corresponding to Eq (1-3) gives

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{M}\mathbf{\Phi} = \mathbf{I} \tag{1-4}$$

$$\mathbf{\Phi}^{\mathrm{T}}\mathbf{K}\mathbf{\Phi} = \mathbf{\Lambda} \tag{1-5}$$

where $\mathbf{\Phi}$ is the modal matrix, $\mathbf{I}$ is the identity matrix, and $\mathbf{\Lambda}$ is the diagonal matrix of the squares of natural frequencies. Then, the physical coordinate vector $\mathbf{x}$ may be represented in the modal space

$$\mathbf{x} = \mathbf{\Phi}\mathbf{q} = \begin{bmatrix} \mathbf{\Phi}_k & \mathbf{\Phi}_d \end{bmatrix} \begin{bmatrix} \mathbf{q}_k \\ \mathbf{q}_d \end{bmatrix} = \mathbf{\Phi}_k\mathbf{q}_k + \mathbf{\Phi}_d\mathbf{q}_d \tag{1-6}$$

where **q** is the modal coordinate vector, and the subscripts *k* and *d* indicate the kept and dropped modes in CMS.

The first several lowest order modes will be kept in the approximation formula in order to represent the physical displacement vector **x**. At the same time, the contribution of those dropped modes in Eq (1-6) still has to be taken into account. For this purpose, let us first assume the modal coordinate vector and force vector take the form

$$\mathbf{q} = \bar{\mathbf{q}}\cos(\omega t) \tag{1-7}$$

$$\mathbf{f}_J = \bar{\mathbf{f}}_J \cos(\omega t) \tag{1-8}$$

where a bar over a symbol indicates the magnitude. Substituting Eqs (1-7) and (1-8) into Eq (1-1), we get

$$\bar{\mathbf{x}} = (\mathbf{K} - \omega^2\mathbf{M})^{-1}\bar{\mathbf{f}} \tag{1-9}$$

When $\omega^2\mathbf{GM}$ is assumed to be small, the last equation may be expanded in matrix Taylor series,

$$\bar{\mathbf{x}} = (\mathbf{I} + \omega^2\mathbf{GM} + \omega^4(\mathbf{GM})^2 + \omega^6(\mathbf{GM})^3 + ...)\mathbf{G}\bar{\mathbf{f}} = \sum_{m=1}^{\infty}\omega^{2(m-1)}(\mathbf{GM})^{m-1}\mathbf{G}\bar{\mathbf{f}} \tag{1-10}$$

where $\mathbf{G} = \mathbf{K}^{-1}$ is referred to as the flexibility matrix.

Eq (1-10) is one of the two expressions of the physical displacement vector **x** that we are interested in. Now, we proceed to acquiring the other one. Let us pre-multiply Eq (1-1) with $\mathbf{\Phi}^T$ and then plug Eq (1-6) into it. We then get a system of decoupled equations

$$\ddot{\mathbf{q}}_i + \omega_i^2\mathbf{q} = \mathbf{g}_i \tag{1-11}$$

where $\mathbf{g} = \mathbf{\Phi}^T\mathbf{f}$, $i = 1, 2, ..., n$. Substituting Eqs (1-7) and (1-8) into Eq (1-11) results in

$$\bar{\mathbf{q}}_i = \frac{\bar{\mathbf{g}}_i}{\omega_i^2 - \omega^2} = \frac{\bar{\mathbf{g}}_i}{\omega_i^2}\left(1 + \frac{\omega^2}{\omega_i^2} + \left(\frac{\omega^2}{\omega_i^2}\right)^2 + \left(\frac{\omega^2}{\omega_i^2}\right)^3 + \ldots\right) = \frac{\bar{\mathbf{g}}_i}{\omega_i^2}\sum_{m=1}^{\infty}\left(\frac{\omega^2}{\omega_i^2}\right)^{m-1} \tag{1-12}$$

where $\frac{1}{\omega_i^2 - \omega^2}$ has been expanded in Taylor series. This is a critical expression whose importance will be seen later. Revisiting Eq (1-6) with Eq (1-7), $\bar{\mathbf{x}}$ can be expressed as

$$\bar{\mathbf{x}} = \bar{\mathbf{x}}_k + \bar{\mathbf{x}}_d \tag{1-13}$$

where

$$\bar{\mathbf{x}}_k = \boldsymbol{\Phi}_k \bar{\mathbf{q}}_k = \sum_{i=1}^{k} \boldsymbol{\varphi}_i \mathbf{q}_i = \sum_{i=1}^{k} \boldsymbol{\varphi}_i \frac{\bar{\mathbf{g}}_i}{\omega_i^2} \sum_{m=1}^{\infty}\left(\frac{\omega^2}{\omega_i^2}\right)^{m-1} = \sum_{m=1}^{\infty} \omega^{2(m-1)} \sum_{i=1}^{k} \boldsymbol{\varphi}_i \frac{\bar{\mathbf{g}}_i}{\left(\omega_i^2\right)^m} \tag{1-14}$$

$$\bar{\mathbf{x}}_d = \boldsymbol{\Phi}_d \bar{\mathbf{q}}_d = \sum_{i=k+1}^{n} \boldsymbol{\varphi}_i \mathbf{q}_i = \sum_{i=k+1}^{n} \boldsymbol{\varphi}_i \frac{\bar{\mathbf{g}}_i}{\omega_i^2} \sum_{m=1}^{\infty}\left(\frac{\omega^2}{\omega_i^2}\right)^{m-1} \tag{1-15}$$

By verification, it is very important to find $\sum_{i=1}^{k} \boldsymbol{\varphi}_i \frac{\bar{\mathbf{g}}_i}{\left(\omega_i^2\right)^m} = \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}^{\mathrm{T}}{}_k \bar{\mathbf{f}}$ in Eq (1-14), and it helps to write Eq (1-13) in a compact form

$$\bar{\mathbf{x}} = \sum_{m=1}^{\infty} \omega^{2(m-1)} \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}_k^{\mathrm{T}} \bar{\mathbf{f}} + \bar{\mathbf{x}}_d \tag{1-16}$$

This is just the second expression of the physical displacement vector $\bar{\mathbf{x}}$ we are pursuing. Combining Eq (1-16) and (1-10), $\bar{\mathbf{x}}_d$ can be written as

$$\bar{\mathbf{x}}_d = \sum_{m=1}^{\infty} \omega^{2(m-1)} [(\mathbf{GM})^{m-1} \mathbf{G} - \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}^{\mathrm{T}}{}_k] \bar{\mathbf{f}} \tag{1-17}$$

Substituting Eq (1-17) into Eq (1-13) gives the desired expression of the physical displacement vector

$$\bar{\mathbf{x}} = \boldsymbol{\Phi}_k \bar{\mathbf{q}}_k + \sum_{m=1}^{\infty} \omega^{2(m-1)} [(\mathbf{GM})^{m-1}\mathbf{G} - \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}^{\mathrm{T}}{}_k] \bar{\mathbf{f}} \tag{1-18}$$

Finally, it is very convenient to define the general *n-th* order approximation formula for the physical displacement vector by truncating higher order terms in Eq (1-18),

$$\bar{\mathbf{x}}^{(n)} = \boldsymbol{\Phi}_k \bar{\mathbf{q}}_k + \sum_{m=1}^{n} \omega^{2(m-1)} [(\mathbf{GM})^{m-1}\mathbf{G} - \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}^{\mathrm{T}}{}_k] \bar{\mathbf{f}} \tag{1-19}$$

where, when $n = 1$, Eq (1-19) gives the expression for $\bar{\mathbf{x}}^{(1)}$ which is the same as the first order approximation, and, when $n = 2$, Eq (1-19) gives the expression for $\bar{\mathbf{x}}^{(2)}$ for the second order approximation, and so on.

It is assumed in above derivation that all substructures are constrained. There are situations where substructures are unconstrained and then have rigid body motion modes. Special procedures are needed to extract elastic modes by removing singularity from stiffness matrix **K** (Craig and Chang, 1977).

2.3 Assembly of substructures with residual flexibility modes

A general free interface component assembly approach that is capable of dealing with complex geometries and configurations was presented by Tournour et al. (2001). For brevity, we consider a structure composed of two substructures that are denoted by α and β .The equations of motion of individual substructure can be explicitly expressed

$$\begin{bmatrix} \mathbf{M}_{II} & \mathbf{M}_{IJ} \\ \mathbf{M}_{JI} & \mathbf{M}_{JJ} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{x}}_I \\ \ddot{\mathbf{x}}_J \end{bmatrix} + \begin{bmatrix} \mathbf{K}_{II} & \mathbf{K}_{IJ} \\ \mathbf{K}_{JI} & \mathbf{K}_{JJ} \end{bmatrix} \begin{bmatrix} \mathbf{x}_I \\ \mathbf{x}_J \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{f}_J \end{bmatrix} \tag{1-20}$$

where subscript I and J indicate interior and interface dofs, respectively. Extraction of equations with respect to the interface dofs from Eq (1-20), application of Eq (1-19) to

the extracted equations, and imposition of the continuity of force and displacement at interfaces

$$\mathbf{f}_J^{(\alpha)} = -\mathbf{f}_J^{(\beta)} \tag{1-21}$$

$$\mathbf{x}_J^{(\alpha)} = \mathbf{x}_J^{(\beta)} \tag{1-22}$$

lead to the general n-th order synthesized equations of motion for the full structure

$$\begin{bmatrix} \boldsymbol{\Lambda}_k^{(\alpha)} + \boldsymbol{\Phi}_{Jk}^{\mathrm{T}\,(\alpha)} \mathbf{K}^{(n)} \boldsymbol{\Phi}_{Jk}^{(\alpha)} & -\boldsymbol{\Phi}_{Jk}^{\mathrm{T}\,(\alpha)} \mathbf{K}^{(n)} \boldsymbol{\Phi}_{Jk}^{(\beta)} \\ -\boldsymbol{\Phi}_{Jk}^{\mathrm{T}\,(\beta)} \mathbf{K}^{(n)} \boldsymbol{\Phi}_{Jk}^{(\alpha)} & \boldsymbol{\Lambda}_k^{(\beta)} + \boldsymbol{\Phi}_{Jk}^{\mathrm{T}\,(\beta)} \mathbf{K}^{(n)} \boldsymbol{\Phi}_{Jk}^{(\beta)} \end{bmatrix} \begin{bmatrix} \mathbf{q}_k^{(\alpha)} \\ \mathbf{q}_k^{(\beta)} \end{bmatrix} = \omega^2 \begin{bmatrix} \mathbf{q}_k^{(\alpha)} \\ \mathbf{q}_k^{(\beta)} \end{bmatrix} \tag{1-23}$$

where

$$\mathbf{K}^{(n)} = \left(\sum_{m=1}^{n} \omega^{2(m-1)} \mathbf{H}_{JJ}^{(m)} \right)^{-1} \tag{1-24}$$

$$\mathbf{H}_{JJ}^{(m)} = \mathbf{G}_{JJ}^{(m)(\alpha)} + \mathbf{G}_{JJ}^{(m)(\beta)}$$

$$\mathbf{G}^{(m)} = (\mathbf{GM})^{m-1} \mathbf{G} - \boldsymbol{\Phi}_k \boldsymbol{\Lambda}_k^{-m} \boldsymbol{\Phi}_k^{\mathrm{T}}$$

Theoretically, Eq (1-23) gives a complete description of general order-reduced synthesized equations of motion for the full structure. However, to solve Eq (1-23) will be impractically difficult when n is given larger than three.

2.4 Numerical validation and application to Monte Carlo Sampling

For illustration purpose, we consider a cantilevered plate, as shown in Fig.1-1, which is divided into two parts. The solutions of the first order and second order free interface CMS are listed in Table 1-1 for in comparison with the ‘exact’ values, i.e., solutions of the full finite element model. It can be seen that there is distinct improvement of accuracy in the second order solutions as compared to the first order solutions.

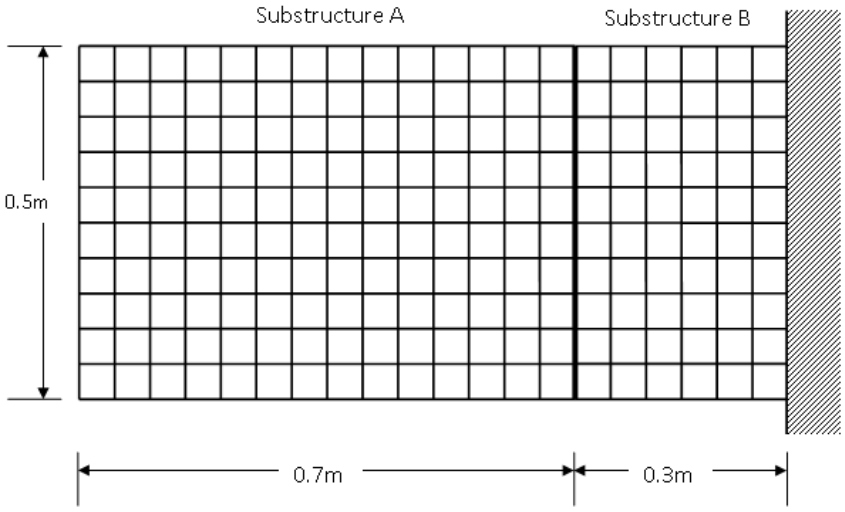

Fig. 1-1 A cantilevered thin plate divided into two substructures

Table 1-1 Eigen-frequencies of the structure: 4 modes retained on each substructure

Mode	First order result	R.E.(%)	Second order result	Relative error (%)	Exact solution
1	107.5591960	0.0276219	107.5295128	0.0000172	107.5294943
2	463.2967982	0.1148072	462.7698774	0.0009437	462.7655103
3	716.8901260	6.9081288	671.9319020	0.2036152	670.5665264
4	1531.5331048	1.6577145	1508.9487806	0.1586474	1506.5586641
5	3303.7301410	75.4222822	2125.4819717	12.8593688	1883.3013108

Though it is difficult to solve such a nonlinear equation of higher order (>=3), in Table 1-1 it is notable that the second-order solutions are very close to ‘exact’ solutions that are obtained from the full model. If quick asymptotical convergence for the second order solutions exists with the increase in the number of kept modes, such solutions may be considered as ‘exact’ values for estimating errors in the first order solutions when

'exact' values are unavailable.

In the following example, a first order CMS model is used for Monte Carlo sampling. It is shown that, when the CMS errors are small enough, using a CMS model to perform Monte Carlo sampling is safe and acceptable. However, to what degree that the CMS errors can be considered "small" is a case-dependent issue. A simplified bladed disk is under investigation, which is composed of one disk and eight blades. Its geometry and finite element mesh are shown in Fig.1-2. The bladed disk is made of homogenous and isotropic steel with constant thickness 0.02m, inner disk radius 0.2m, outer disk radius 0.6m, and blade length 0.4m. The material density, Yong's modulus, and Poisson ratio are 7870kg/m^3, $2.10\times10^{11}\ \text{N/m}^2$, and 0.3, respectively. The finite element model of the structure is built using the thin plane theory. Note the geometry is complex and thus the assembly of the CMS model is carried out using the procedure presented by Tournour et al. (2001). There are 1920 degrees of freedom on the whole integral structure. In CMS modeling, only the first lowest 5 modes of each component are retained and so the reduced model has only 45 degrees of freedom. The two sets of solutions obtained by the full model and the CMS model are listed in Table 1-2 for comparison. The solutions given by the CMS model show that, although only the first-order approximation is adopted, the relative errors in the natural frequencies of the first a few lowest modes are very small.

To introduce uncertainties in the structure, it is assumed that the length of each blade is subject to normally distributed variation with the mean value 0.4m and standard deviation 0.002m. Monte Carlo simulations then are performed using the full finite

element model and the reduced model separately. In Fig.1-3 histograms of eigenvalues for mode 1, mode 4, and mode 7 are shown for comparison. As we can see, the differences between the results of the full finite model and the CMS model are little enough to be neglected. However, not all CMS models can give such satisfactory results as the CMS model in this simple case does. This is particularly true with those structures having complex geometry and complicated uncertainties. In Chapter 4, two-level Gaussian process regression is presented to address this issue.

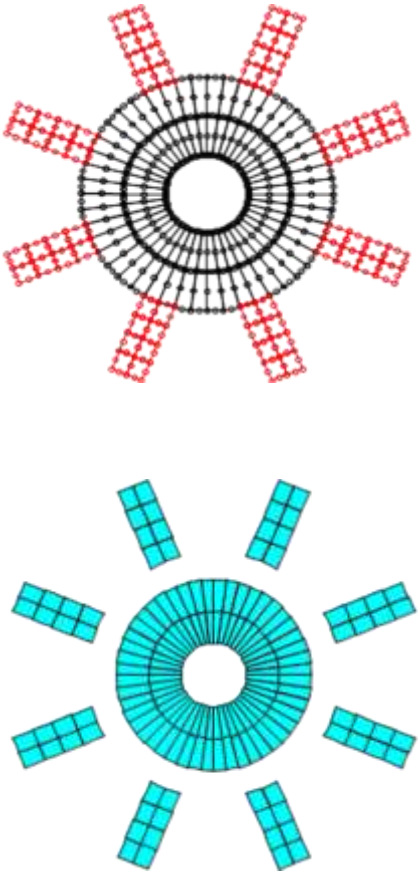

Fig. 1-2 A simplified bladed disk. Top – Geometry and meshing of the structure; Bottom - Decomposition of the structure for CMS modeling

Table 1-2 Comparison of natural frequencies: the full FE model v.s. the CMS model

Mode	FEM (Hz)	CMS (Hz)	Relative Error
1	41.3404872	41.3404879	0.0000017%
2	41.3404872	41.3404879	0.0000016%
3	41.7063471	41.7063477	0.0000016%
4	45.2321765	45.2321801	0.0000081%
5	45.2321771	45.2321801	0.0000067%
6	55.6507692	55.6579795	0.0129563%
7	55.6507693	55.6579795	0.0129561%
8	60.7779083	60.7795137	0.0026413%

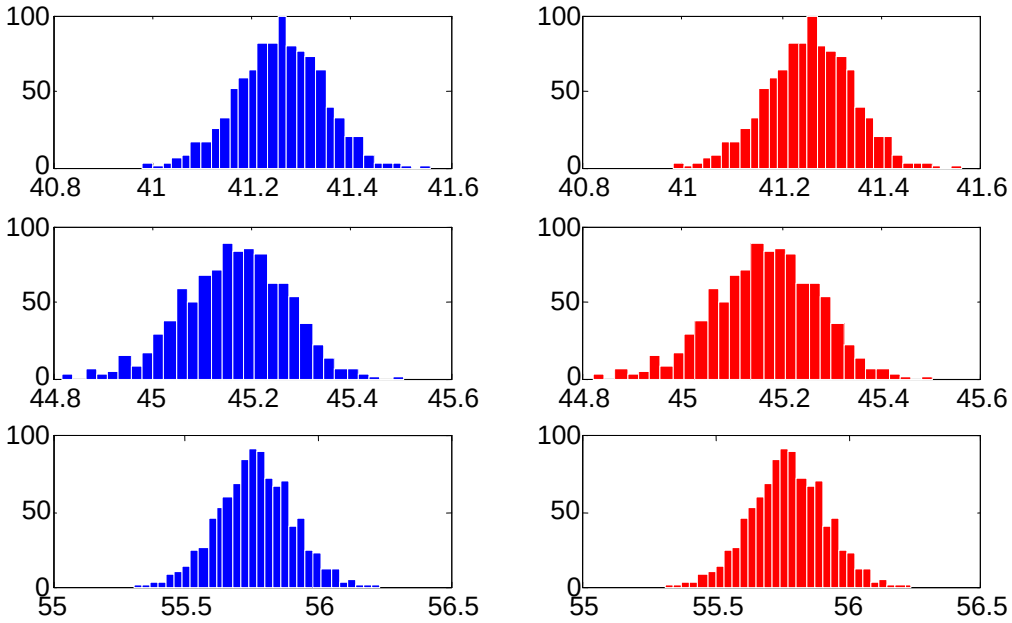

Fig. 1-3 Histograms of natural frequency: direct FEM (left) v.s. CMS (right)
Vertical axis: class frequency; Horizontal axis: frequency (Hz)
Top: mode 1; Middle: mode 4; Bottom: mode 7

Chapter 3: Gaussian Process Regression

3.1 Introduction

Characterization of dynamic behaviors plays an important role in structural design, assessment, and control, especially in cases where uncertainties on structures may lead to significant abnormalities in dynamic responses. On a structure, there are various sources of uncertainty such as material properties, manufacturing tolerance, and in-service degradation. Uncertainties generally appear to be random phenomena. Some of them such as material property variations may be modeled in terms of probabilities, and others such as geometry variations are difficult to be quantified. There are two basic approaches to introducing uncertainties into structural dynamic analyses. The parametric approach identifies model parameters individually and treats them as random variables within the structural models. The non-parametric approach considers the overall effects of all the uncertainties on the structural dynamical behaviors without specific analyses of individual uncertainties. As far as finite element based approaches are concerned, the representative parametric methods include stochastic finite element analysis and perturbation methods (Ghanem and Spanos, 1991; Panayirci et al., 2011; Friswell and Adhikari, 2000), and a typical non-parametric method is proposed by Soize (2000) to model uncertainties based on the random matrix theory. It is noticed that all these methods are trying to incorporate probabilistic models of uncertainties into structural finite element modeling frameworks. Furthermore, all of them eventually require data to quantify parameters in their probabilistic models of uncertainties, even in the non-parametric approach (Soize, 2008; Adhikari and Sarkar, 2009). This is due to the nature

of uncertainty analysis which essentially deals with random variables and fundamentally has to rely on data or observations. Therefore, data and sampling techniques are two indispensible and closely related factors in performing uncertainty analysis. Compared with stochastic finite element analysis and perturbation methods, Monte Carlo uncertainty analysis is more straightforward. It directly samples data through repeatedly running a deterministic model at different parameter configurations. Its results are also more robust because it requires no additional formulations to combine probabilistic concepts and structural finite element modeling. Such formulations always lead to inaccuracy because more mathematical approximations or physical simplifications have to be involved.

In modern engineering practice, however, structural dynamic analyses often require large-scale finite element models, especially when analyzing complex structures, and running of them may be computationally expensive. The feasibility of Monte Carlo uncertainty analysis thus becomes questionable when these expensive models are used for sampling. Some researchers have introduced order-reduced models to Monte Carlo uncertainty analysis of structural dynamics (Hinke et al., 2009; Bladh et al., 2002; Lee and Castanier, 2006; Dohnal et al., 2009; Panayirci et al., 2011). Initially order reduction techniques were developed to reduce the computational costs of deterministic finite element analyses in structural dynamics (Tournour et al., 2001; Craig, 2000; Feiner and Griffin, 2002; Moyroud et al., 2002; Mbaye et al., 2010; Bladh et al., 2001; Bah et al., 2003).Order reduction techniques use only the most relevant information extracted from a full finite element analysis to reconstruct a small-sized approximate model, which may

lead to the similar robustness issue of the non-Monte-Carlo methods mentioned above. Xia and Tang (2011) presented a two-level Gaussian process regression method to address the effects of the approximations employed in order reduction techniques on the accuracy of Monte Carlo quantification of structural responses under uncertainty.

To ensure the fidelity of data, many practical applications require full finite element models to be used for deterministic analyses but also for Monte Carlo sampling. Because of the high computational costs involved, however, the information obtained from the deterministic analyses may be very limited and the resulting samples also may be very small. Under this data availability constraint, Gaussian processes can be introduced to structural dynamic analysis, especially Monte Carlo uncertainty analysis. The basic idea behind Gaussian processes is to extend the discrete multivariate Gaussian distribution on a finite-dimensional space to a random continuous function defined on an infinite-dimensional space. Gaussian processes have been well developed over the last decade. The main contribution to this achievement came from the continual efforts in the field of machine learning (Neal, 1998; Mackay, 2003; Rasmussen and Williams, 2006) as well as the ever-increasing data processing capabilities of modern computers. In the field of structural dynamic analysis, DiazDelaO and Adhikari (2010) pioneered the use of Gaussian processes to emulate the frequency responses in a deterministic analysis of a simple plate structure. They also attempted to use Gaussian processes to emulate the realizations of random fields and the implementations of polynomial chaos expansions in stochastic finite element analysis (DiazDelaO and Adhikari, 2011). It is well recognized that stochastic finite element theory is derived rigorously from the probability theory but

has a profound numerical difficulty of incorporating random processes in the abstract domain of a probability space (Ghanem and Spanos, 1991). On the other hand, Gaussian process regression is an inference method that is primarily based on observed data other than physical mechanisms of systems. Generally, if the underlying mechanism of a system is unknown or poorly understood, inference methods such as Gaussian processes are more practical and offer more plausible predictions about the behaviors of the system. It is also in the most recent years that Gaussian process regression methods have been introduced to other engineering areas such as system controls and nonlinear system modeling (Hasen et al., 2005; Gregorcic and Lightbody, 2009; Azman and Kocijan, 2009; Mohanty et al., 2009).

The purpose of this chapter is to present Gaussian process regression as an efficient approach to analyzing structural dynamics, especially characterizing structural responses under uncertainty. For this purpose, three inference schemes are specially designed to show that there are different ways to use Gaussian processes to address specific needs in structural frequency response analyses. These extended applications include inferring special representative quantities of frequency responses that summarize critical structural dynamic properties of complex structures with uncertainty. Additionally, an important feature of the Gaussian process regression method introduced in the present thesis is the employment of the Bayesian approach to estimate the parameters in the regression model, which greatly enhances the learning ability of Gaussian processes. Considering the difficulty of integrals in a full Bayesian approach, a flexible two-stage hierarchical Bayesian framework is then adopted to accommodate the more practical maximum

likelihood method to estimate the parameters associated with covariance functions (Kennedy and O'Hagan, 2000a). Furthermore, it should be clear that in our discussions a system generally may refer to a real physical system or process, a physical model, or a numerical model. Gaussian processes can be applied to any other kind of systems though a finite element model is used in this thesis for illustration purposes.

The remaining parts of this chapter are organized as follows: Chapter 3.2 introduces the basics about Bayesian approach which provides a fundamental framework to estimate parameters in Gaussian process regression models; Chapter 3.3 is an outline of the Gaussian process regression with special details on the hierarchical Bayesian approach for separately estimating parameters and hyper-parameters; Chapter 3.4 presents three Gaussian process application examples for structural dynamic analysis using a particularly chosen cyclically symmetric structure; and the concluding remarks are given in Chapter 3.5.

3.2 Bayesian parameter estimation

The Bayesian probability theory evolved from the underlying idea of the Bayesian theorem, and superseded the traditional frequency definition of probability. Theoretically, the Bayesian approach is a perfectly formulated procedure that strictly follows the inductive logic to formalize our knowledge based on both our beliefs and observations under circumstances of uncertainty (Sivia and Skilling, 2006). The details on the Bayesian probability and statistics can be readily found in literature (Loredo, 1990; Mockus, 1997; Sivia and Skilling, 2006). In this section we give a concise description of

the Bayesian parameter estimation which is used to determine the parameters and the hyper-parameters in the Gaussian process regression method introduced in the present paper.

Let a system be represented by a function $y = f(\mathbf{x})$, where $\mathbf{x} \in R^m$ is an m-dimensional input vector and $y \in R$ is a scalar output. The system output may be a scalar or a vector variable. This paper limits discussions to the scalar output case because multiple output Gaussian process regression still needs more research in the areas of statistics and machine learning (Alvarez and Lawrence, 2011). Given a set of n observations, so-called training data in Gaussian processes, which is denoted as $\mathbf{D} = \{(y_i, \mathbf{x}_i) \mid i = 1, 2, ..., n\}$ where each $(y_i, \mathbf{x}_i)$ is referred to as a data point, the goal is to make prediction for the function output $y*$ at a given target point $\mathbf{x}*$. We assume that the data points are conditionally independently distributed so that the likelihood, which is a function that returns the probability of the data set **D** conditional on given parameters, is calculated by multiplication of the probabilities of all individual data points

$$P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{n} P(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}) \tag{2-1}$$

where $\boldsymbol{\theta}$ is the set of system parameters and $P(y_i \mid \mathbf{x}_i, \boldsymbol{\theta})$ is the probability of each data point conditional on $\boldsymbol{\theta}$. Note $\boldsymbol{\theta}$ may be parameters of a physical system or a model. As we see in the ensuing sections, $\boldsymbol{\theta}$ is the collection of both parameters and hyper-parameters in a Gaussian process regression model. The Bayesian inference requires a prior probability $P(\boldsymbol{\theta})$ to represent our beliefs, or prior knowledge, about the parameters before data points are observed. Given the observed data points, our knowledge about $\boldsymbol{\theta}$

is updated by applying the Bayesian theorem, which gives birth to the posterior probability distribution of $\boldsymbol{\theta}$

$$P(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) = \frac{P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}) P(\boldsymbol{\theta})}{P(\mathbf{y} \mid \mathbf{X})} \tag{3-2}$$

where $\mathbf{X} = [\mathbf{x}_{1,} \mathbf{x}_{2,}, ..., \mathbf{x}_n]^{\mathrm{T}}$ and $\mathbf{y} = [y_{1,} y_{2,}, ..., y_n]^{\mathrm{T}}$ are the notations for the set of the training input vectors and the set of the observed outputs, respectively, and the denominator is the marginal likelihood of the parameters

$$P(\mathbf{y} \mid \mathbf{X}) = \int P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{3-3}$$

which is a normalization constant that may be omitted in this research since model selection is not under consideration (Sivia and Skilling, 2006; Rasmussen and Williams, 2006). With the updated probability distribution of the parameters $\boldsymbol{\theta}$, the posterior predictions for target cases $(\mathbf{y}^*, \mathbf{X}^*)$ are evaluated by averaging all possible outputs on the parameter space

$$P(\mathbf{y}^* \mid \mathbf{X}^*, \mathbf{y}, \mathbf{X}) = \int P(\mathbf{y}^* \mid \mathbf{X}^*, \boldsymbol{\theta}) P(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) d\boldsymbol{\theta} \tag{3-4}$$

Essentially, this is a calculation of the weighted mean across the parameter space. Notice the presence of integrals in Eqs (3-3) and (3-4), which in practical applications are usually intractable analytically and require numerical approximation methods (Rasmussen and Williams, 2006).

As seen in Eq (3-2) that our beliefs about the parameters $\boldsymbol{\theta}$ are represented by $P(\boldsymbol{\theta})$ and are modified through the likelihood $P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\theta})$ in the light of the observed date points, the Bayesian inference offers an evolutionary process to make plausible

propositions in cases where it is not possible to argue with certainty (Sivia and Skilling, 2006). We emphasize that the Bayesian approach is a *process* because it is an open-ended way whereby a user can update both the prior knowledge and the conclusions when more information is obtained from new observations.

3.3 Gaussian process regression

In this section, a brief description of the Gaussian process regression within a hierarchical Bayesian framework is presented, and important implementation details are discussed specifically.

3.3.1 Gaussian process regression within Bayesian framework

Gaussian processes are a data-based approach by which various spatial and temporal problems can be modeled. There has been a rich research literature on regression using Gaussian processes and the readers are encouraged to refer to the works by Mackay (2003) and Rasmussen and Williams (2006) for complete details. As defined in the preceding section, a system is represented by a function $y = f(\mathbf{x})$ and a data set is denoted as **D**. Given the data points **D**, to predict the target cases $(\mathbf{y}^*, \mathbf{X}^*)$ is straightforward if, for any set of input points $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]^T$, the distribution of the corresponding observed outputs $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ is assumed to be a multivariate Gaussian. The multivariate Gaussian distribution of a discrete subset of the range of the function can be extended to its entire continuous range. Thus, the multivariate Gaussian distribution for a finite-dimensional case is generalized to an infinite-dimensional case.

The function $y = f(\mathbf{x})$ with the generalized Gaussian distribution of infinite dimensionality over its domain is referred to as a Gaussian process. The Gaussian assumption can be justified by the central limit theorem in many applications, especially in cases where multiple independent uncertainty sources are present.

A Gaussian process is fully specified by a mean function $m(\mathbf{x}_i)$ and a covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$. The mean function $m(\mathbf{x}_i) = E\left(f(\mathbf{x}_i)\right)$ is an assumed function that returns the expected value of $f(\mathbf{x}_i)$. The covariance function $k(\mathbf{x}_i, \mathbf{x}_j) = E\left((f(\mathbf{x}_i) - m(\mathbf{x}_i))(f(\mathbf{x}_j) - m(\mathbf{x}_j))\right)$ is an assumed function to relate the observations at $\mathbf{x}_i$ and $\mathbf{x}_j$. For a data set $\mathbf{D} = \{(y_i, \mathbf{x}_i) \mid i = 1, 2, ..., n\}$, the n-dimensional expected value vector and the $n \times n$ covariance matrix of the multivariate Gaussian distribution are constructed using the assumed mean function $m(\mathbf{x}_i)$ and covariance function $k(\mathbf{x}_i, \mathbf{x}_j)$. Thus, a multivariate Gaussian distribution defined on a space of discrete vectors is generalized to a Gaussian process on a space of continuous functions, which is symbolically expressed as

$$f \sim \mathbf{GP}(m, k) \tag{3-5}$$

Given that the observed data are subject to error or noise, it is necessary to distinguish between the observed output, denoted as y, and the true output of the system, denoted as $f(\mathbf{x})$. Their relationship is assumed to be $y = f(\mathbf{x}) + \varepsilon$, where ε is referred to as the noise term, representing the errors in numerical results or the noise in experimental data. The data set **D** is presented in the standard form of the multivariate Gaussian distribution

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{3-6}$$

where $\mathcal{N}$ stands for Gaussian distribution, and the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are defined respectively as

$$\boldsymbol{\mu} = \begin{bmatrix} m(\mathbf{x}_1) \\ m(\mathbf{x}_2) \\ \vdots \\ m(\mathbf{x}_n) \end{bmatrix} \tag{3-7}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{2n} & \cdots & k_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{bmatrix} \tag{3-8}$$

Thorough discussions on the selection of mean and covariance functions can be found in the works by Mackay (2003) and Rasmussen and Williams (2006). In our research a linear model is adopted for the mean function

$$m(\mathbf{x}) = h(\mathbf{x})\boldsymbol{\beta} \tag{3-9}$$

and the covariance function is represented by the combination of a squared exponential and an independent identically distributed Gaussian

$$k_{ij} = \sigma_f^2 \exp\left\{-b(\mathbf{x}_i - \mathbf{x}_j)^T(\mathbf{x}_i - \mathbf{x}_j)\right\} + \sigma_n^2 \delta_{ij} \tag{3-10}$$

where $h(\mathbf{x})$ is a base function vector to be determined in the subsequent discussions, δ_{ij} is the Kronecker delta, $\sigma_n^2 \delta_{ij}$ represents the noise terms, $\boldsymbol{\beta}$ is the vector gathering the parameters associated with the mean function, and $\phi = \{\sigma_f, b, \sigma_n\}$ is the set of the parameters associated with the covariance function, referred to as the hyper-parameters in

literature. We can see that $\sigma_f^2 + \sigma_n^2$ is the allowable maximum covariance value and b is a roughness factor that regulates the influence of the distance between two observations on the covariance value and thus decides the smoothness of the curve of predicted outputs.

Since the fundamental assumption is that any set of the outputs of the system follows the multivariate Gaussian distribution defined in Eq (3-6), the predictions of $\mathbf{y}*$ at target input points $\mathbf{X}*$ can be presented together with the observed data points in the standard form

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}* \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}* \\ \boldsymbol{\Sigma}*^{\mathbf{T}} & \boldsymbol{\Sigma}** \end{bmatrix} \right) \tag{3-11}$$

where $\boldsymbol{\mu}*$ is the mean vector with respect to $\mathbf{x}*$, $\boldsymbol{\Sigma}*$ is the sub-matrix containing $k(x_i, x_j*)$, and $\boldsymbol{\Sigma}**$ is the sub-matrix containing $k(x_i*, x_j*)$. It can be shown that the probability distribution of $\mathbf{y}*$ at the target input points $\mathbf{X}*$ conditional on the data set **D** is also a multivariate Gaussian (Rasmussen and Williams, 2006)

$$\mathbf{y}* \mid \mathbf{X}*, \mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \phi \sim \mathcal{N}\left(\boldsymbol{\mu}* + \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \boldsymbol{\Sigma}** - \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}* \right) \tag{3-12}$$

The expected value of $\mathbf{y}*$ is $\boldsymbol{\mu}* + \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and the corresponding covariance matrix is $\boldsymbol{\Sigma}** - \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}*$ where the main diagonal elements are the variance values and the off-diagonal elements are the covariance values.

As discussed in Chapter 3.2, given the data set **D**, the Bayesian approach is used to update the parameters and the hyper-parameters defined in Eq (3-9) and (3-10). To that end, a two-stage hierarchical Bayesian framework is adopted to estimate the parameters

and the hyper-parameters separately (Kennedy and O'Hagan, 2000a). In the high stage, conditional on both the posterior distribution of the hyper-parameters ϕ and the data set **D,** the distribution of the posterior parameters $\boldsymbol{\beta}$ is given by the Bayesian theorem

$$P(\boldsymbol{\beta} \mid \mathbf{y}, \mathbf{X}, \phi) = \frac{P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \phi) P(\boldsymbol{\beta} \mid \phi)}{P(\mathbf{y} \mid \mathbf{X}, \phi)} \tag{3-13}$$

where the denominator is the marginal likelihood

$$P(\mathbf{y} \mid \mathbf{X}, \phi) = \int P(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \phi) P(\boldsymbol{\beta} \mid \phi) d\boldsymbol{\beta} \tag{3-14}$$

In the low stage conditional on the data set **D** the distribution of the posterior hyper-parameters ϕ is also given by the Bayesian theorem

$$P(\phi \mid \mathbf{y}, \mathbf{X}) = \frac{P(\mathbf{y} \mid \mathbf{X}, \phi) P(\phi)}{P(\mathbf{y} \mid \mathbf{X})} \tag{3-15}$$

with the marginal likelihood

$$P(\mathbf{y} \mid \mathbf{X}) = \int P(\mathbf{y} \mid \mathbf{X}, \phi) P(\phi) d\phi \tag{3-16}$$

The evaluations of the posterior distribution of the hyper-parameters ϕ and the posterior distribution of the parameters $\boldsymbol{\beta}$ are carried out sequentially from the low stage to the high stage. The posterior mean hyper-parameters $\hat{\phi}$ and the posterior mean parameters $\hat{\boldsymbol{\beta}}$ are then determined accordingly. This is the so-called process of training the regression model. Finally, the posterior predictions about target cases $(\mathbf{y}^*, \mathbf{X}^*)$ can be obtained by simply substituting $\hat{\phi}$ and $\hat{\boldsymbol{\beta}}$ into Eq (3-12). Due to the difficulty of evaluating the integrals in the full Bayesian approach as described above, the maximum likelihood method is proposed to estimate the posterior hyper-parameters $\hat{\phi}$ in the low

stage of the two-stage Bayesian framework (Kennedy and O'Hagan, 2000a, 2000b). It is therefore a semi-Bayesian approach (Swiler, 2006) since the use of the observed data **D** in the maximum likelihood method to estimate the posterior hyper-parameters $\hat{\phi}$ violates the underlying assumption in the Bayesian approach that the prior distribution of ϕ solely depends on its own parameters but not on the data **D**.

3.3.2 Implementation details

The implementation of Gaussian process regression involves first the acquisition of data points, the selection of forms of the mean and covariance functions, and the assignment of priors to the parameters and the hyper-parameters. The selection of training input points can be done by experimental design (DiazDelaO and Adhikari, 2010; Kennedy and O'Hagan, 2000a). In this research we randomly draw input points from the parameter configuration space so as to test the robustness of the Gaussian process regression. The training outputs corresponding to the selected training inputs are obtained either by running a numerical code or by performing physical experiments. These training input and output points constitute the training data set **D** as defined in Chapter 3.2. The selection of the forms of mean and covariance functions for the Gaussian process regression model is founded on reasonable assumptions about the underlying relationship between the observed inputs and outputs of the system. The works by Mackay (2003) and Rasmussen and Williams (2006) have provided a rich collection of mean and covariance functions. The mean and covariance functions that are used in this research have been described in Chapter 3.1.

The key step is then to determine the posterior hyper-parameters $\hat{\phi}$ and the posterior parameters $\hat{\boldsymbol{\beta}}$. We first need to designate a set of prior probability distributions for ϕ and $\boldsymbol{\beta}$ to best present our prior knowledge. In this research we take the suggestion by Kennedy and O'Hagan (2000a, 2000b, 2001). A set of independent non-informative priors are chosen for $\boldsymbol{\beta}$, σ_f^2, and b such that $P(\boldsymbol{\beta},\sigma_f^2,b)\propto b^{-1}\sigma_f^{-2}$ and σ_n is a zero-mean Gaussian $\sigma_n \sim \mathcal{N}(0,\sigma_n^2)$. The full Bayesian approach for estimating the hyper-parameters is numerically intensive. Within the two-stage hierarchical Bayesian framework described above, the posterior hyper-parameters $\hat{\phi}$ are estimated using the maximum likelihood method, which is relatively easy to conduct but requires a choice made from multiple optimal solutions. Given the posterior hyper-parameters $\hat{\phi}$, the posterior parameters $\hat{\boldsymbol{\beta}}$ are obtained in a closed form (Kennedy and O'Hagan 2000b), $\hat{\boldsymbol{\beta}} = (\mathbf{h}(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{h}(\mathbf{X}))^{-1}\mathbf{h}(\mathbf{X})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$, where $\mathbf{h}(\mathbf{X})$ is the notation for $\left[h(\mathbf{x}_1) \quad h(\mathbf{x}_2) \quad \cdots \quad h(\mathbf{x}_n)\right]^{\mathrm{T}}$. With the posteriors $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$ and the data set $\mathbf{D}$, the posterior predictions $\mathbf{y}^*$ at target input points $\mathbf{X}^*$ are evaluated by directly using Eq (3-12). Further implementation details are specifically given in the following numerical examples.

3.4 Structural dynamic analysis using Gaussian processes

Gaussian process regression methods use data to predict structural responses.This is fundamentally different from the conventional physical mechanism based methods. One

of the major advantages of the regression methods is the reduction in the costs for analysis, especially for uncertainty analysis where extensive sampling is often required. In this section, we use a simplified bladed disk to illustrate the applications of Gaussian processes to structural dynamic analyses. This particular structure belongs to the class of cyclically symmetric structures that exhibit high sensitivity with respect to uncertainties. Dynamic analyses of this class of structures in engineering practice typically demand complex modeling and vast computational resources.

3.4.1 Modeling: a simplified bladed disk

We investigate a bladed disk as seen in Fig.3-1, which is a simplified representation of the industrial engine bladed disks. In reality this type of structures has the spatial quasi-cyclic symmetry and is highly sensitive to the uncertainties among its blades. Fig.3-1 shows the geometry and meshing scheme of the bladed disk. It comprises one disk and 18 blades, with constant thickness 0.005m, inner disk radius 0.13m, outer disk radius 0.19m, and blade length 0.12m. The structure is supposed to be made of homogenous isotropic steel and the nominal mass density, Young's modulus, and Poisson ratio are $7870\ \mathrm{kg/m^3}$, $2.10\times10^{11}\ \mathrm{N/m^2}$ and 0.3, respectively. The inner boundary is fixed. The finite element model is built with 360 thin plate elements and totally has 4104 degrees of freedom. The equations of motion can be expressed as

$$\mathbf{M}\ddot{\mathbf{x}}+\mathbf{C}\dot{\mathbf{x}}+\mathbf{K}\mathbf{x}=\mathbf{F} \tag{3-17}$$

where **M**, **C**, and **K** are the mass, damping and stiffness matrices, respectively, **x** is the displacement vector, and **F** is the external force vector. The damping matrix **C** is modeled

as weak proportional modal damping with a constant ratio $\zeta = 0.001$. It is supposed that the bladed disk is subject to the engine order excitation. Specifically, the engine order excitation is applied on the 18 blades and the excitation configuration on each blade is given as $F_j = F_0 e^{i(\omega t + \phi_j)}$ ($j = 1, \ldots, N$; N is the number of the blades), where F_j is the external forces applied on the *j*-th blade, F_0 is the force magnitude, ω is the excitation frequency, and $\phi_j = [2\pi C(j-1)]/N$ (C is the engine order number, $C = 0,1,\cdots,\text{int}(N/2)$) is the phase shift on the *j*-th blade.

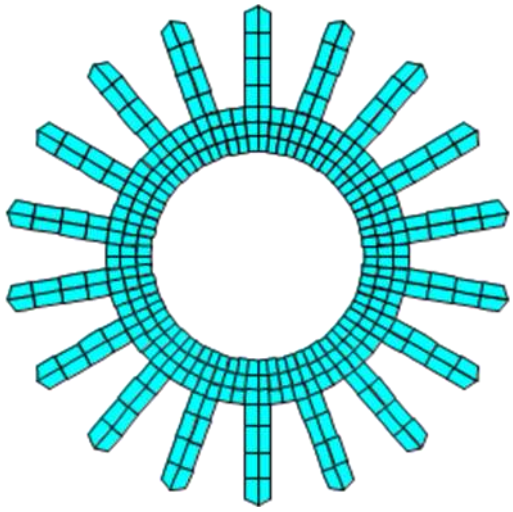

Fig. 3-1 Geometry and meshing of the example bladed disk

The ideal situation is that a tuned bladed disk has the perfect cyclic symmetry and thus all the blades are identical to each other. The perfect cyclic symmetry groups the vibration modes of the tuned bladed disk into different mode families. A mode family contains a set of closely spaced natural frequencies and the associated mode shapes that

exhibit distinct spatial harmonic patterns. In Table 3-1 the second column is a list of the natural frequencies of the second mode family of the tuned bladed disk. We see that all the modes except the first and the last ones within a mode family are paired and each pair of modes has the same natural frequencies. When the external excitation of a specific engine order number is applied within the frequency range of a mode family, the very vibration mode within the family that has exactly the same spatial harmonic index as the engine order number will dominate the responses of the tuned bladed disk.

When analyzing the mistuning problem of a bladed disk, it is usually supposed that the disk is perfectly cyclically symmetric but the blades differ from each other due to small variations in geometry or material properties. Basically uncertainties such as these small variations can be introduced to structural dynamics analyses in a parametric or a non-parametric fashion. In this research it is assumed that the 18 blades have different Young's moduli while the Young' modulus within an individual blade is uniform. The 18 varying Young's moduli, $\mathbf{E} = [E_1 \quad E_2 \quad \cdots \quad E_N]^{\mathrm{T}}$ ($N = 18$), thus constitute a 18-dimensioanl random parameter configuration space which is modeled as a multivariate Gaussian $\mathbf{E} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In Table 3-1 the third column lists a set of the natural frequencies of a mistuned bladed disk where the mistuning is introduced by a parameter vector $\mathbf{E} = [E_1 \quad E_2 \quad \cdots \quad E_N]^{\mathrm{T}}$ which is randomly sampled from the random space $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with a 5% standard deviation (SD). In the third column, all the previously paired modes that are shown in the second column no longer appear in pairs due to the mistuning among the blades. An associated phenomenon is that some blades may possess much larger modal amplitudes than others, which is the so-called localization of vibration

modes and has been investigated extensively in turbomachinery (Whitehead, 1998; Kenyon and Griffin, 2003).

Table 3-1 Natural frequencies (Hz) of the second mode family, mode 19 to 36

Mode	Tuned	Mistuned (SD = 5%)
19	873.16	870.34
20	877.76	875.92
21	877.76	881.24
22	893.22	893.00
23	893.22	898.80
24	923.75	924.59
25	923.75	926.63
26	973.57	972.44
27	973.57	977.56
28	1043.00	1042.48
29	1043.00	1048.55
30	1125.41	1127.06
31	1125.41	1128.03
32	1207.31	1204.99
33	1207.31	1213.58
34	1270.61	1269.31
35	1270.61	1279.54
36	1295.29	1299.47

In the following examples, the engine order excitation is designed as an array of external forces within the specified frequency range exerted at the tips of the 18 blades. The magnitude of each force is one unit. The frequency responses of the bladed disk are measured also at the tips of the 18 blades and are represented in the form of a vector of N ($N = 18$) elements, each of which is the frequency response amplitude u_i at the tip of the i-th blade

$$\mathbf{u}(\boldsymbol{\omega}, \mathbf{E}) \equiv \begin{bmatrix} u_1(\boldsymbol{\omega}, \mathbf{E}) \\ u_2(\boldsymbol{\omega}, \mathbf{E}) \\ \vdots \\ u_N(\boldsymbol{\omega}, \mathbf{E}) \end{bmatrix} \tag{3-18}$$

where $\boldsymbol{\omega}$ is the vector of the frequencies swept within the frequency range of interest and $\mathbf{E}$ is a specific 18-dimensional Young's modulus configuration. Fig.3-2(a – d) present a series of comparisons of the frequency responses under the second engine order excitation between the tuned bladed disk and four different mistuned bladed disks. The frequency responses of the tuned bladed disk are depicted with solid lines. All the tuned responses are identical and only one resonant peak appears around the second natural frequency of the mode family as listed in Table 3-1. The dotted lines represent the frequency responses of the mistuned bladed disks. The mistuned frequency response analyses are performed at four parameter configurations that have different standard deviations of Young's modulus, i.e., 2%, 5%, 8%, and 12%. In each subplot, it is evident that the patterns of the mistuned response curves of the 18 blade tips seem to be varying 'randomly' as compared with the uniformity of the tuned response curves. Given different degrees of mistuning, the irregularities of the mistuned responses can be quantified by two distinctive features: the split distance of the response peaks and the change of the maximum peak value. Comparing the mistuned responses in the four subplots in Fig.3-2, it can be seen that the increase of mistuning, which is realized by increasing the standard deviation of Young's modulus from 2% to 12% in this example, generally leads to a larger response peak value as well as a broader split of response peaks. Since the frequency response curves of mistuned bladed disks apparently look like

a set of realizations of a random function on a sample space, Monte Carlo simulation is employed to characterize these varying responses across the parameter configuration space.

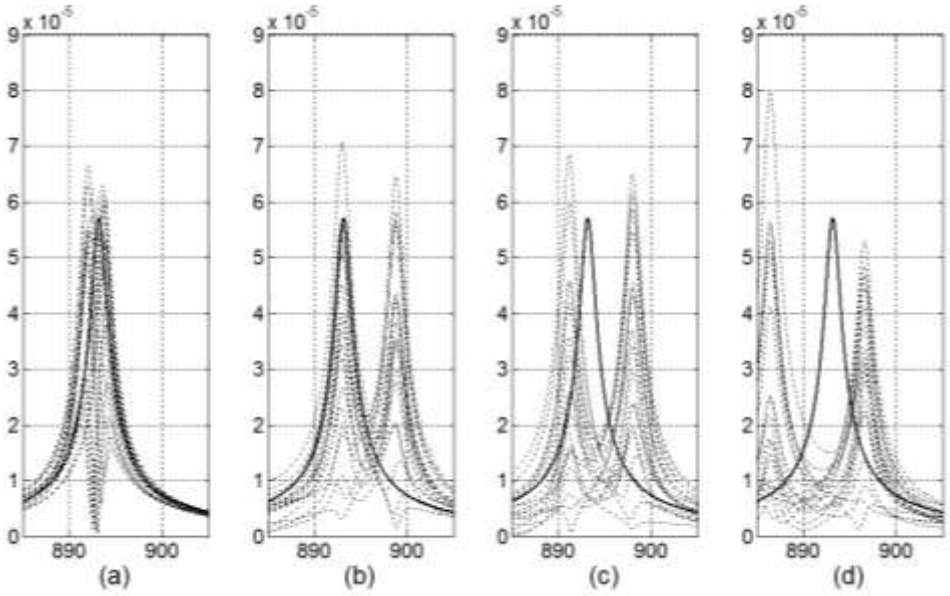

Fig. 3-2 The 2nd engine order frequency responses on 18 blade tips: horizontal axis - frequency (Hz); vertical axis - amplitude (m); solid lines – tuned responses; dotted lines - mistuned responses.
(a) - 2% mistuning standard deviation; (b) - 5% mistuning standard deviation; (c) - 8% mistuning standard deviation; (d) - 12% mistuning standard

The standard Monte Carlo procedure requires that a set of random variables are drawn from a defined sample space and then a model is run repeatedly to produce outputs corresponding to the sampled variables. The resulting Monte Carlo sample is required to be large enough to cover the major collective response characteristics of the structure under uncertainty. Fig.3-3 depicts a Monte Carlo sample of the frequency responses that are measured at the tip of the second blade. It is acquired by repeatedly running the deterministic finite element model of the bladed disk at 1000 parameter configurations

that are sampled from the aforementioned 18-dimensional random space $\mathbf{E} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with a 5% standard deviation using the Latin hypercube sampling method. The maximum response peak value, denoted as crosses in Fig.3-3, of each individual response curve within the frequency range is a quantity of special interest in cases where potentially dangerous excessive responses of the blades are considered. However, it is time-consuming and resource-demanding to run the finite element frequency response simulator at different parameter configurations. Even a single run of the simulator at only one parameter configuration may be extremely costly if the high resolution frequency sweep is required to fully expose the important response features. In the following discussions, therefore, the first application example is to use Gaussian processes to predict frequency responses within the entire frequency range of interest by using sampled responses at a few sampled frequencies. Then two more complicated examples are presented to show that brute-force Monte Carlo simulations can be replaced by Gaussian processes to make inferences about special representative quantities, such as response envelops and maximum peaks, derived from the frequency responses within the given frequency range on the parameter configuration space.

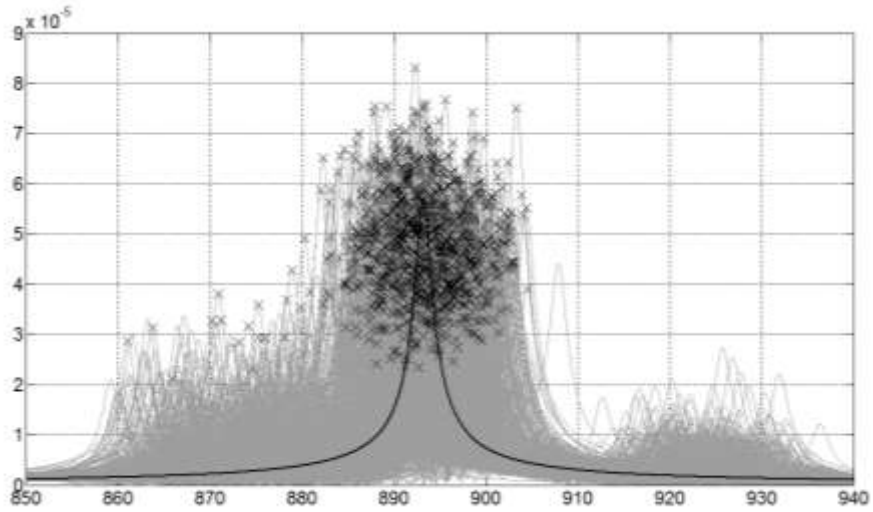

Fig. 3-3 The 2nd engine order frequency responses on blade 2 by Monte Carlo Sampling (SD = 5%): horizontal axis - frequency (Hz); vertical axis – amplitude (m); grey lines – 1000 sampled responses; black line – nominal responses; crosses – response peaks of individual runs

3.4.2 Frequency response prediction

In this first example, Gaussian process regression is applied to predicting the frequency responses of the bladed disk under engine order excitations, which is essentially similar to the example presented by DiazDelaO and Adhikari (2010). The numerical evaluations of the frequency responses using finite element models at specific parameter configurations are procedurally straightforward as described in the last sub-section.The computational costs, however, may be prohibitively high due to the possible large model sizes. This is particularly true with cyclically symmetric structures such as bladed disks that have clustered natural frequencies and thus demand high resolution frequency sweeps to evaluate the frequency responses. Similar issue exists for experimental frequency response analyses of complex structures. Gaussian process

regression as a data-based statistical tool is thus employed to make predictions about frequency responses of the bladed disk. This inference approach requires that only the frequency responses at a small discrete set of frequencies sampled from the entire frequency range of interest are evaluated with the finite element model and then are used as the training data for the Gaussian process regression model to predict the frequency response at target frequencies. As compared with the number of frequency points required by a full frequency sweep, the number of the sampled frequency points is very small, and therefore the data acquisition costs are significantly lowered. Given the sample frequency responses as training data, moreover, running a Gaussian process regression model is also relatively cheap. Thus, a large amount of computational costs can be saved for Gaussian processes to predict the frequency responses within the entire frequency range based on the sparsely sampled frequency responses.

For illustration purposes, the quantity of interest in this example is the frequency response at the tip of the second blade of a mistuned bladed disk. The Gaussian process emulation of the frequency responses within the entire frequency range $\boldsymbol{\omega}$ can be expressed as a function that approximates the related finite element frequency response simulator

$$u_2(\boldsymbol{\omega}) \approx g(\boldsymbol{\omega}, \boldsymbol{\beta}, \phi) \tag{3-19}$$

where the Young's modulus configuration $\mathbf{E}$ of the bladed disk is a constant vector randomly sampled from the 18-dimensional random space $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This mistuned bladed disk is a deterministic system, and this example shows the application of Gaussian processes to deterministic analyses. To work out the frequency responses within the

specified frequency range from 885 Hz to 905 Hz as plotted in Fig.3-4, we first need to acquire a set of data points. We select, or more precisely, sample, say 9, frequencies as the training input points, denoted as **X** in Eq (3-12), and then evaluate, by running the finite element frequency response simulator, the corresponding frequency responses as the training output points, denoted as **y** in Eq (3-12). Each training input point and its corresponding training output point form a data point, as denoted as plus sign in Fig.3-4, and these data points constitute a training data set, generically called data, used for training purposes in Gaussian processes. Specifically, training a Gaussian process regression model is the process of updating the pre-assumed hyper-parameters ϕ and parameters $\boldsymbol{\beta}$ in the model using the information extracted from the data points. Note that in some application situations the selection of the training input points is determined by design (Kennedy and O'Hagan, 2000a). Fig.3-4(a), (d), and (g) depict the frequency responses predicted by a Gaussian process regression model that has been trained using the 9 data points. Although exactly the same 9 data points are used in Fig.3-4(a), (d) and (g), the predictions are definitely different in terms of the smoothness of the mean response curves and the widths of the credible intervals. Fig.3-4(a) has the bumpiest mean response curve and broadest credible intervals, Fig.3-4(d) has the smoothest mean response curve and moderate interval widths, and Fig.3-4(g) has a moderately smooth mean response curve and the narrowest credible intervals.

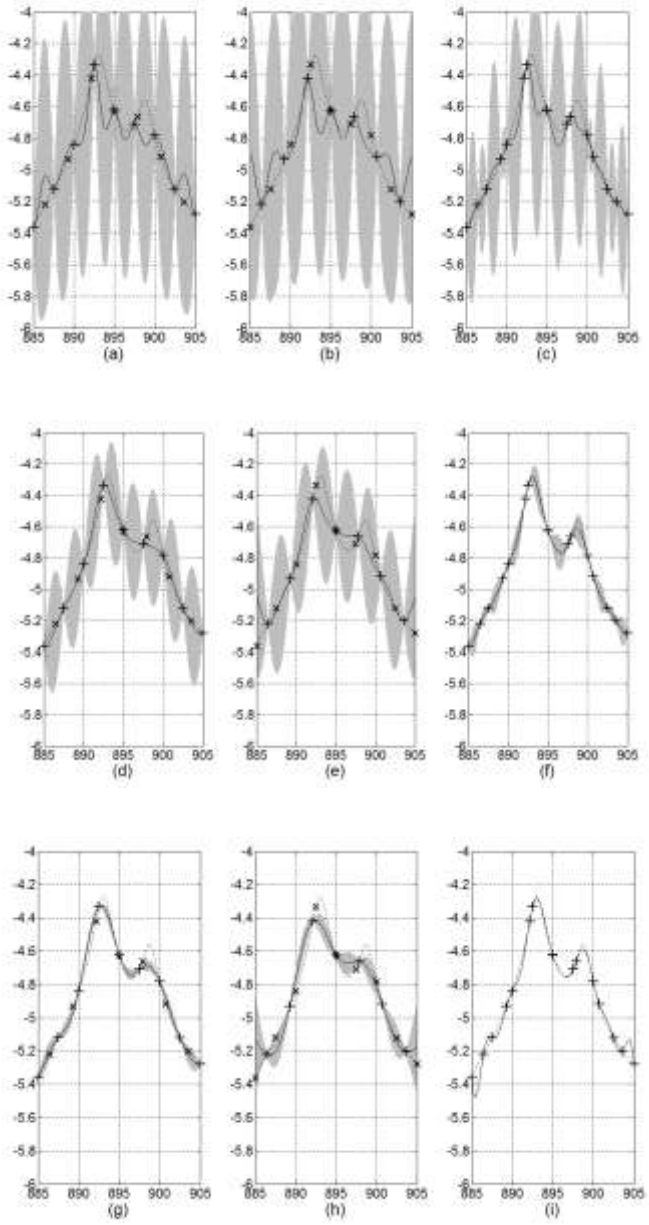

Fig. 3-4 The 2nd engine order frequency responses at the tip of blade 2: horizontal axis - frequency (Hz); vertical axis - log amplitude; plus – training point; cross – validation point; solid line – inferred responses; dotted line – true responses; shaded area – 95% credible interval; Plots (a), (d), and (g) are obtained with same 9 training data but different initial hyper-parameters; Plots (b), (e), and (h) are obtained with same 7 training data but different initial hyper-parameters; Plots (c), (f), and (i) are obtained with different posterior hyper-parameters $\hat{\phi} = \{1.0, 0.1996, 0\}$ **(top),** $\{0.3, 0.1039, 0\}$ **(middle), and** $\{0.07, 0.1134, 0\}$ **(bottom).**

What leads to the different predictions as seen in Fig.3-4(a), (d), and (g) reflects a challenging issue in the implementation of statistical modeling and analysis. In the Bayesian analysis the issue is about how to determine the convergence of the Markov chain Monte Carlo (MCMC) algorithm used in evaluating the posterior predictions (Brooks, 1998); in the maximum likelihood approach it is about how to find the global optimal solution as we often only find a set of local optimal solutions that are substantially influenced by initial values. As discussed in Chapter 3-3, the full Bayesian approach to evaluating the posterior parameters and the posterior hyper-parameters requires approximation methods to conduct the integrals as formulated in Eqs (3-14) and (3-16), respectively. A widely used class of approximation algorithms is the MCMC method. However, the MCMC method typically is numerically expensive, which may offset the efficiency gained by the Gaussian process regression method. In our research, the full Bayesian approach is bypassed by using the maximum likelihood method, which is a typical practice for the implementation of Gaussian process regression, especially when a relatively rich set of training data are available (Kennedy and O'Hagan, 2000a; Rasmussen and Williams, 2006). Particularly, within the two-stage hierarchical Bayesian framework introduced in Section 3, as the noise term σ_n in Eq (3-10) is modeled as a Gaussian, the logarithm likelihood function of the hyper-parameters ϕ is in a closed form (Kennedy and O'Hagan, 2000a, 2000b)

$$L = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}[\mathbf{y}-\mathbf{h}(\mathbf{X})\boldsymbol{\beta}]^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}[\mathbf{y}-\mathbf{h}(\mathbf{X})\boldsymbol{\beta}] \tag{3-20}$$

where $\mathbf{h}(\mathbf{X}) = \left[h(\mathbf{x}_1) \quad h(\mathbf{x}_2) \quad \cdots \quad h(\mathbf{x}_n)\right]^{\mathrm{T}}$ in which n is the number of data points. For instance, in the 9-data-point case described above, the input points are sampled frequencies $\mathbf{X} = \boldsymbol{\omega}$ where $\boldsymbol{\omega} = [\omega_1 \quad \omega_2 \quad \cdots \quad \omega_9]^{\mathrm{T}}$ and correspondingly $\mathbf{h}(\boldsymbol{\omega}) = \left[h(\omega_1) \quad h(\omega_2) \quad \cdots \quad h(\omega_9)\right]^{\mathrm{T}}$. Basically it is a trial-and-error process to select the base functions $h(\omega)$ for the mean function $m(\omega)$. In doing so, one should keep in mind that, amongst a collection of candidate models, it is not guaranteed that a sophisticated model shall be better than a simple one in terms of the universality of models (MacKay, 2003). In our case it is found that the choice $h(\omega_i) = [\omega_i \quad 1]$ is good enough to give satisfactory results. In the low stage of the two-stage hierarchical Bayesian framework, the posterior hyper-parameters $\hat{\phi}$ are estimated by maximizing the logarithm likelihood defined in Eq (3-20) and then used as fixed values in Eq (3-13) to calculate the posterior parameters $\hat{\boldsymbol{\beta}}$. With the posteriors $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$, the posterior predictions $\mathbf{y}^*$ at the target input points $\mathbf{X}^*$ are evaluated directly by using Eq (3-10).

The multiple optimal solutions encountered in using the maximum likelihood method to determine the posterior hyper-parameters $\hat{\phi}$ are shown in Fig.3-4(a), (d), and (g), where the differences are due to different initial values assigned to the hyper-parameters ϕ. The likelihood of the hyper-parameters ϕ given the data set **D** usually is multimodal, namely multiple local optima can be found in the process of optimization. A practical issue thus arises on how to decide the "best" optimal solution among a set of local optimal solutions that are sensitive to the initial values of the hyper-parameters ϕ

assigned to the optimization algorithm. Generally, the multiplicity of local optima may be relieved or eliminated by choosing more informative priors and constructing a more sophisticated regression model (Mackay, 2003; Rasmussen and Williams, 2006; Sivia and Skilling, 2006). It continues to be an active research area of Bayesian analysis and mathematical optimization. In the following discussions, we take a heuristic cross-validation approach to assessing different optimal solutions for the 'best' choice of the posterior hyper-parameters $\hat{\phi}$ (Rasmussen and Williams, 2006).

Besides the 9 data points as mentioned above, we select 7 more data points and thus have a data set that includes two subsets: the original 9 data points and the additional 7 data points. The two subsets are used for training and validation purposes alternately. In Fig.3-4(a), (b), (d), (e), (g), and (h), the training data points and validation data points are denoted as pluses and crosses, respectively. The first step is to train the Gaussian process regression model using the subset of the 9 data points and validate the resulting predictions using the subset of the 7 data points, which are depicted in Fig.3-4(a), (d), and (g). The second step is to reverse the use of the two subsets of data points by using the subset of the 7 data points to train the model and the subset of the 9 data points to validate the results, which are depicted in Fig.3-4(b), (e), and (h). We then compare the validating results with respect to different initial hyper-parameters ϕ in the two steps and choose the posterior hyper-parameters $\hat{\phi}$ that make the Gaussian process regression model best represent all the 16 data points. In this illustrative example the comparison can be carried out simply by visual inspection. The results in Fig.3-4(a-b) have the broadest credible intervals, indicating that the initial hyper-parameters ϕ lead to a

regression model that overestimates the uncertainties in its predictions. By contrast, the results in Fig.3-4(g-h) have the narrowest credible intervals that fail to cover some validation data points, which indicates the initial hyper-parameters ϕ lead to a regression model that underestimates the uncertainties in its predictions. The initial hyper-parameters ϕ in Fig.3-4(d-e) lead to a regression model that gives the predictions not only best fitting all the 16 data points but also covering all important response features, especially the two peaks, with reasonable credible intervals. At this point a knowledgeable judgment has to be made based on the user's experiences to choose the "best" posterior hyper-parameters $\hat{\phi}$ from the three optimal solutions.

In Fig.3-4(c), (f), and (i), the posterior hyper-parameters $\hat{\phi}$ are, respectively, $\{1.0, 0.1996, 0\}$, $\{0.3, 0.1039, 0\}$, and $\{0.07, 0.1134, 0\}$, which are the arithmetic average of the posterior hyper-parameters obtained from the preceding two steps of alternate training and validation. The value of the roughness factor *b* has different orders of magnitude in the three sets of hyper-parameters $\hat{\phi}$, which implies that the roughness factor *b* plays a dominant role as compared with σ_f^2. Because the system discussed in this example is the computer code of a deterministic finite element model of the bladed disk that returns definite outputs without uncertainty, the hyper-parameter for the noise term, σ_n^2, which is designed to model uncertainty in outputs of the system, is set to be zero. In the case of dealing with a system of which the observed outputs involves noises or uncertainty, the noise term should be turned on in the Gaussian process regression model.

It is worth mentioning that in this example we have chosen a very small modal damping ratio, $\zeta = 0.001$, in order to yield a model of the bladed disk that has high sensitivity in its frequency responses with respect to mistuning. As shown in Fig.3-2 and 3, the frequency response curves of this slightly damped bladed disk drastically fluctuates with sharp resonant peaks. If the damping is even smaller, more training data have to be acquired for training the Gaussian process regression model so that its predictions cover more features of the frequency response curves, especially those around the natural frequencies of the structure. However, considering the cost increase due to more data points required for accurate emulation of the frequency response, in some cases it is necessary to re-select a quantity of interest that not only contains the essential information to meet the goal of analysis but also can be measured more conveniently. For instance, if the goal of analysis is to avoid prescribed excessive responses, it is more efficient to directly predict the response peaks on the parameter space than to emulate all the related response curves individually. In the following examples Gaussian processes are used to make inferences about two types of critical representative quantities derived from frequency responses to address specific concerns in structural analyses.

3.4.3 Extreme response prediction under uncertainty

In the preceding sub-section, the Gaussian process regression is used to make inferences for the frequency responses at a particular location on the bladed disk at a fixed parameter configuration.The general procedure and the implementation specifics of the Gaussian process regression are also illustrated. In the context of uncertainty analysis,

we are interested in the response variations of the structure on the random parameter configuration space. In this sub-section, we introduce a Gaussian process inference scheme to predict the response envelope that contains the possible maxima of the frequency responses of the bladed disk within the frequency range of interest.

Conventionally, a large Monte Carlo sample has to be generated to construct such a response envelope. That is, within the specified frequency range we need to numerically calculate or experimentally measure the frequency responses at all the sampled parameter configurations and find the maximum response at each frequency. Fig.3-5(a) presents an example where a set of response curves are acquired from 1000 runs of the finite element frequency response simulator. The upper envelope, denoted as the dotted line, is the collection of the maxima of all the sampled responses at individual frequencies within the whole frequency range. This upper envelope is referred to as the 'true' upper envelope to serve as the reference for assessing the inferred envelope by Gaussian processes. Such an upper envelope, summarizing the possible risky extreme responses due to mistuning within a bladed disk, is important for assessing its reliability and durability. Obviously, if a single run of the finite element model or the experimental testing is expensive, the brute-force Monte Carlo approach as demonstrated in Fig.3-5(a) is impractical.

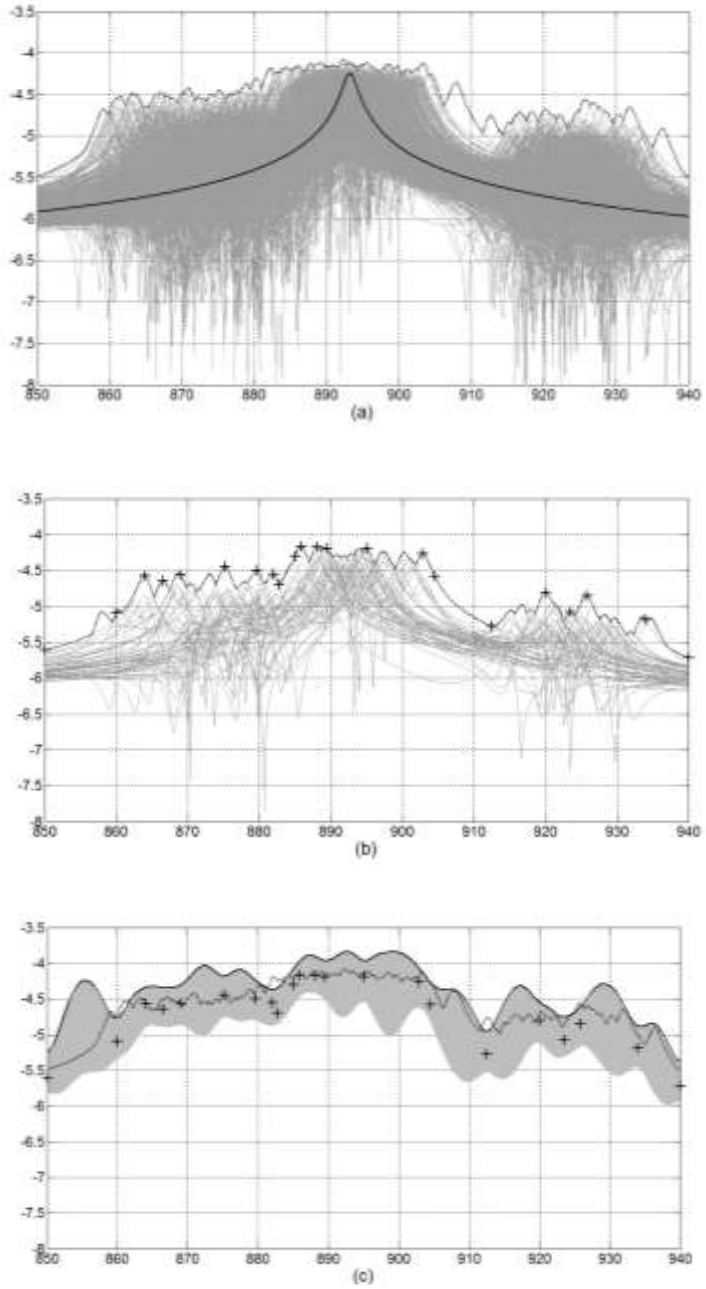

Fig. 3-5 Statistical characterization of frequency responses on blade 2: (a) - Monte Carlo sampling with 1000 runs; (b) - Monte Carlo sampling with 50 runs; (c) - Inferred envelope compared with true envelope; Horizontal axis - frequency (Hz); Vertical axis - log amplitude; Grey line in (a) and (b) - individual sampled response curve; Dotted line in (a) and (c) - upper envelope of 1000 sampled response curves; Dotted line in (b) - upper envelope of 50 sampled response curves; Solid line in (a) - Nominal response curve; Solid line in (c) - Inferred upper envelope; Plus sign - feature point; Shaded area in (c)- 95% credible interval

Let us suppose that only a relatively small Monte Carlo sample is available, e.g., 50 sample points in this numerical example. All the 50 sampled response curves and the corresponding upper envelop are plotted in Fig.3-5(b). We are only interested in the possible maximum response at each frequency, and we hope to infer an upper envelope to contain all possible maximum responses within the specified frequency range on the parameter configuration space. Particularly, we expect all possible response peaks fall beneath the inferred upper envelope. To that end, we first identify and select a set of feature points in the upper envelope that is constructed using the 50 sampled response curves. As seen in Fig.3-5(b), the plus signs indicate the feature points selected to represent the major characteristics of the upper envelop. By ignoring all the minor details of the envelope, we are trying to improve the robustness of our predictions. We use the Gaussian process regression model with the selected feature points as the training data to predict the mean values and the 95% credible intervals within the specified frequency range. Note that the interpretation of a credible interval at a frequency is the probability that the true value of the frequency response of the system happen to be within that interval at that frequency. Therefore, as depicted in Fig.3-5(c), the upper bounds of the 95% credible intervals based on the sampled response curves are considered to be used to constitute the inferred envelop to quantify all possible maximum responses within the specified frequency range.

The procedure of working out the inferred envelope is similar to that of working out the inferred frequency responses in the last sub-section, but there are two important differences. First, these feature points in the upper envelope of the sampled frequency

responses are carefully selected as the training data, as opposed to the intentionally randomly sampled training points used in the example of the last sub-subsection. Second, the upper envelop can never be known with certainty and can only be inferred plausibly based on the observed maximum responses at individual frequencies within the specified frequency range. As seen in Fig.3-5(c), the posterior hyper-parameters are given as $\hat{\phi} = \{0.07, 0.1063, 0.0110\}$, where the noise term is turned on to determine the uncertainty in the predicted frequency response maxima. In Fig.3-5(c), the inferred envelope based on the 50 sampled frequency response curves is plotted against the 'true' envelope as described above. It can be seen that the inferred envelope successfully contains all the response peaks as well as almost the entire 'true' envelop. This example suggests that, in cases where the availability of numerical realizations or experimental measurements are very limited, the Gaussian process inference scheme that takes advantage of the interpretation of the credible intervals provides a practical approach to predicting possible extreme values of quantity of interest.

3.4 .4 Emulation for Monte Carlo sampling

There are three categories of variables in a system: parameters, inputs, and outputs. Chapter 3.4.1 shows how to perform emulations of the outputs of the system corresponding to the inputs, and Chapter 3.4.2 shows how to make inferences about the outputs of the system on the random parameter configuration space. In this sub-section, we present an inference scheme that directly models the underlying relationship between the output space and the parameter configuration space of the system. Gaussian process

emulation is directly used for Monte Carlo sampling on the parameter configuration space. Notwithstanding the possible loss in accuracy, an emulator is desirable if it can generate a sufficiently large sample at a relatively low cost to convey major statistical characteristics of the outputs of a random system. As stressed in the Introduction, the emulation scheme illustrated in this example can be applied to either numerical or physical data on the condition that the parameter configuration space is well defined for the system.

The outputs of interest is the response peak values at the tip of the second blade of the bladed disk on the random 18-dimensional Young's modulus configuration space within the frequency range from 850 Hz to 940 Hz under the second engine order excitation. Let $f(\mathbf{E})$ denote the finite element model of the bladed disk that returns the frequency response peak value at the tip of the second blade within the specified frequency range at the given parameter configuration **E**. The goal is to develop a Gaussian process regression model $g(\mathbf{E}, \boldsymbol{\beta}, \phi)$ to approximate $f(\mathbf{E})$

$$f(\mathbf{E}) \approx g(\mathbf{E}, \boldsymbol{\beta}, \phi) \tag{3-21}$$

The linear model $m(\mathbf{E}) = h(\mathbf{E})\boldsymbol{\beta} = \sum_{j=1}^{18} E_j \beta_j + \beta_0$ is selected as the mean function that represents our belief that the random vector **E** is varying within the close neighborhood of its nominal value on the parameter configuration space and the peak response values vary in an approximately linear way correspondingly.

The procedure for the emulation is the same as that in Chapter 3.4.1. Note that the input space of the Gaussian process regression model is now the random 18-dimensional

Young's modulus configuration space of the finite element model. From the 1000 response peaks in the example depicted in Fig.3-3, we randomly draw 21 data points to train the Gaussian process regression model. This example deals with a scalar function $f(\mathbf{E})$ that has an 18-dimensional input parameter vector containing the 18 Young's moduli of the 18 blades. The ideal way to assess the predictions against the true values is to plot them on a 19-dimensional space. It is, however, impossible to plot such a high-dimensional space. We therefore use 2-dimenisonal plots as shown in Fig.3-6 and 7, where lines are used to connect the predicted output points which actually are scattered on the 19-dimensional space. To some degree, the connecting lines in Fig.3-6 and 7 depicts the underlying trends among data on the 19-dimensional space only in a symbolical sense, but really help to visually assess the match between the predictive regression model $g(\mathbf{E},\boldsymbol{\beta},\phi)$ and the finite element model $f(\mathbf{E})$. In Fig.3-6(a) the inferences about the peak response values at the remaining 979 parameter configurations are plotted to compare with the 'true' values, i.e., the peak values given by the finite element model. The distribution of the inferred peak values shifts a little downward as a whole. Nevertheless, the maxima of the inferred response peak values are always higher than those of the true response peak values. This is desirable because it indicates that the range of the peak values predicted by the regression model can contain all true values, especially those extreme true values. In Fig.3-6(b) we can see clearly that the trend line of the inferred values is roughly parallel to that of the true values, which suggests that, to a considerable degree, our regression model has captured the underlying relationship

between the space of the random 18-dimensional parameter configurations and the space of the peak response values returned by the finite element model.

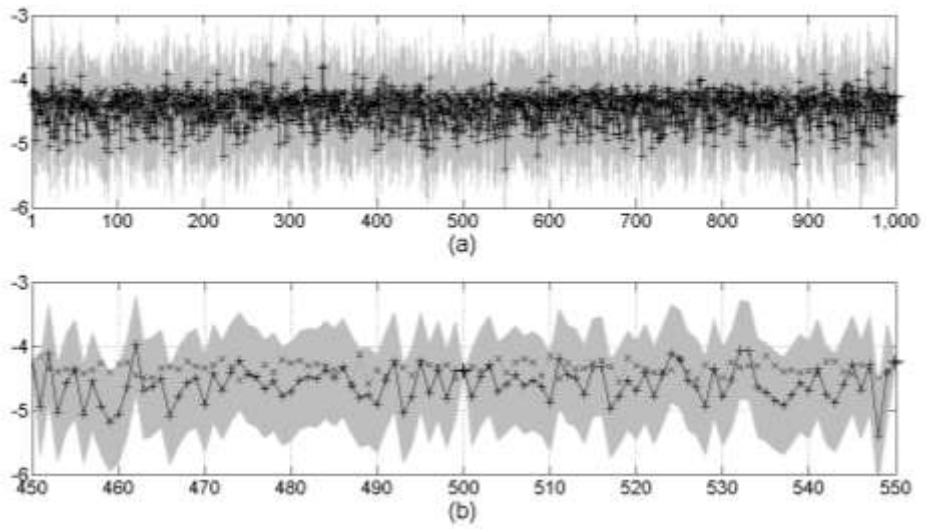

Fig. 3-6 Response peaks inferred with 21 training data (a) at all 1000 configuration points; (b) in a zoomed-in region: horizontal axis – sample point number; vertical axis – log amplitude; cross – true peak value; plus – inferred peak value; bold plus – data point; shaded area – 95% credible interval

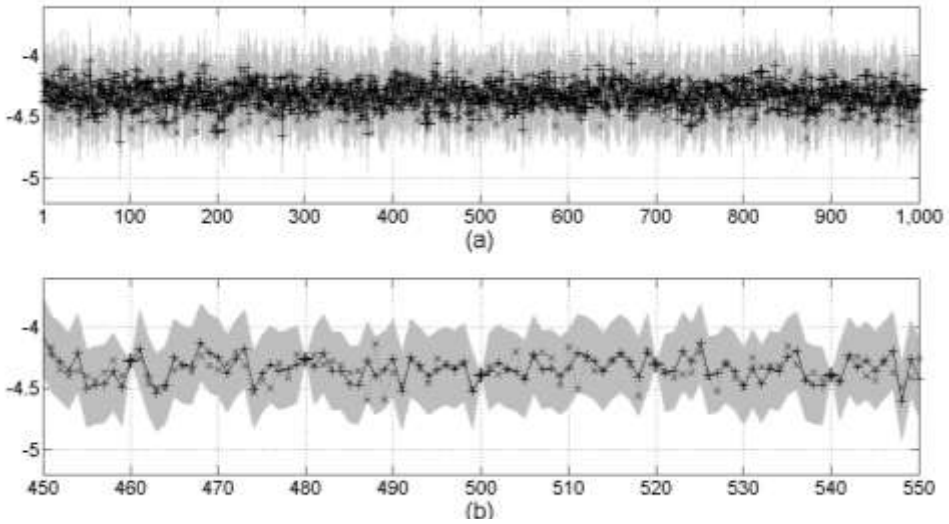

Fig. 3-7 Response peaks inferred with 51 training data (a) at all 1000 configuration points; (b) in a zoomed-in region: horizontal axis – sample point number; vertical axis – log amplitude; cross – true peak value; plus – inferred peak value; bold plus – data point; shaded area – 95% credible interval

The first attempt to improve the accuracy of the inferences is to use more training data. Fig.3-7 depicts the case where 51 data points are drawn from the same 1000 sample points to train the Gaussian process regression model. From Fig.3-7(a) it is evident that the inferred response peak values at the remaining 949 parameter configurations are closer to the true values than those in Fig.3-6(a). There is no downward shift of the distribution of the inferred response peak values. Note that in both the two cases in Fig.3-6 and 7 the data points are chosen randomly, and a better way is to pre-select data points by design so as to avoid the distribution shift in Fig.3-6. In the present paper we intentionally choose data points randomly in order to demonstrate the flexibility and robustness of the Gaussian process regression. The improvement is obvious when comparing the results in Fig.3-6(b) with those in Fig.3-7(b). In the latter case, the discrepancy between the inferred value and the true value at each individual parameter configuration is much smaller. The trend line of the inferred response peak values is also closer to that of the true response peak values.

To measure how well the inferred response peak values match the corresponding true response peak values, we scatter the two sets of values in a plane where the horizontal and vertical coordinates stand for, respectively, the inferred values and the true value, as shown in Fig.3-8. Each dot point in the plane denotes the match between an inferred value and its corresponding true value. The crosses that are distributed in the diagonal line indicate the perfect matches. Fig.3-8(a) and (b) present the matches for the two cases in Fig.3-6 and 7, respectively. In Fig.3-8(a) and (b), the two sets of dot points are distributed roughly in two elliptical shapes with different sizes and axial directions. The

perfect match line divides each of the distributions into two halves. In Fig.3-8(a) the major axis of the distribution is not aligned with the perfect match line and there are more points in the left half, which reflects the downward shit of the distribution of the inferred response peak values in Fig.3-6. In Fig.3-8(b) there are more date points used for training the regression model. The results are significantly improved in two senses.First, the distribution of the inferred values is more concentrated around the perfect match line.Second, its major axis is well aligned with the perfect match line, evenly dividing the distribution into two nearly identical halves. Notice the outliers in the right half of the distribution are of particular interest because it indicates that the inferred peak values can cover the maximum of the true peak values. Comparisons among Fig.3-6, 7, and 8 demonstrate that simply increasing data points may lead to the improvement of the accuracy of inference results of the Gaussian process regression. A thorough validation analysis of the Gaussian process emulation is presented in the paper by Bastos and O'Hagan (2009). The fact that the inference accuracy improves with the increase of data points demonstrates the learning ability of the Gaussian process regression model. This is an important property of Gaussian processes that better performance can be achieved by extracting more information form additional data points without having to modify the model structure, e.g., re-selecting the mean function and the covariance function.

So far all the examples presented above have clearly shown that the Gaussian process regression can be an efficient analyzing tool to address the data availability constraints. As far as the computational time is concerned, the Gaussian process emulation generally is much faster than the brute-force Monte Carlo simulation. In the example of this sub-

section, for instance, the total computational time required by an emulation includes the time of running the finite element model for acquiring data points and the time of running the Gaussian process regression model. On a Dell® T7500 workstation it takes about 285 seconds to run the finite element model to generate a data point. The 1000 Monte Carlo sample points consume about 79.17 hours. In the case where 51 data points are used, it takes 4.04 hours for data acquisition but only 33 seconds for the Gaussian process regression model to work out the inferred response peak values at the remaining 949 sampled parameter configurations. Obviously the acquisition of data points accounts for almost all the time required by the emulation. Comparing the 79.17 hours consumed for acquiring the 1000 Monte Carlo sample points with the 4.04 hours consumed for acquiring the 51 data points to train the regression model, we see that the use of the Gaussian process emulation leads to a dramatic reduction in the computational time.

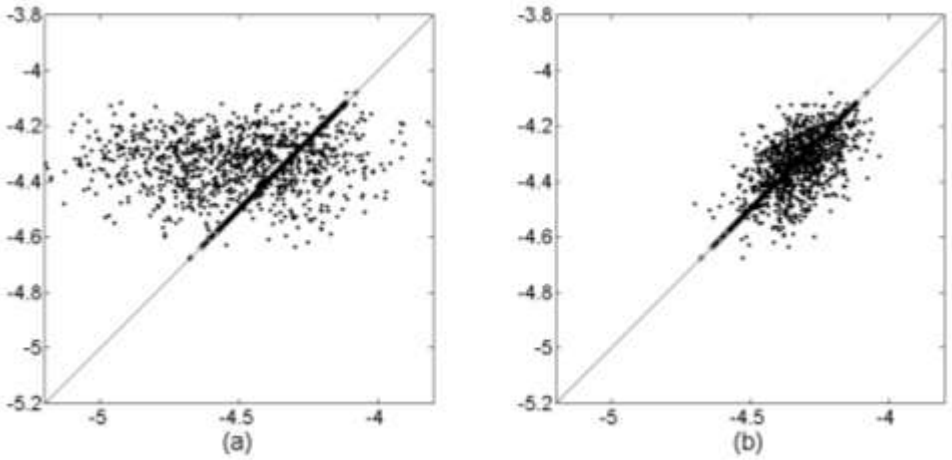

Fig. 3-8 Comparison of true and inferred peak values (a) with 21 training data; (b) with 51 training data: horizontal axis – inferred value; vertical axis – true value; plus – true peak; dot – inferred peak

3.5 Concluding remarks

In this chapter we present a complete brief description of Gaussian processes within the Bayesian framework and propose a series of schemes to make inferences about frequency responses and associated critical representative quantities for structures with uncertainty. A simplified bladed disk is employed for case study. The effectiveness and efficiency of the Gaussian process regression approach for characterizing structural dynamic responses is demonstrated through numerical examples, from the simple emulations of frequency responses to the high-dimensional regression modeling and prediction with uncertainty. In conjunction with Monte Carlo sampling, Gaussian process regression models use only relatively small samples to make predictions about the collective response characteristics and computational costs thus can be saved remarkably. A two-stage hierarchical modeling strategy has been adopted for Gaussian processes to bypass the difficulty in numerical integration involved in a full Bayesian framework. The examples also suggest that, since we typically have no full knowledge about a system under investigation, it is important to use heuristic methods, such as cross-validation and user's judgments, to address practical issues encountered in implementations. For this purpose, we give a thorough discussion on the 'best' choice of the posterior hyper-parameters for the covariance function in a Gaussian process regression model.

Chapter 4: Two-Level Gaussian Regression

4.1 Introduction

In this chapter we are going to explore the combination of CMS modeling and Gaussian process emulation to realize rapid Monte Carlo sampling.Questions arise regarding how to assess a sample obtained by an emulator, especially when training data are given by a CMS model, and, if possible, how to lessen the inevitable errors in the sample. Clearly, errors exist in the results of both the CMS model and the Gaussian process emulator. The errors in the solutions given by CMS models, called CMS errors, mainly come from the truncation of higher order normal modes of substructures as well as the approximation for compensating modes. When a CMS model is used for Monte Carlo sampling, CMS errors distort statistical measures such as mean and standard deviation of the sample. The errors in the results of Gaussian process emulation are called inference errors due to the nature of inference itself. Because of the learning ability of Gaussian processes, inference errors can be reduced by simply suing more data points, which has been demonstrated by Xia and Tang (2011). However, when CMS results are used as training data, CMS errors are inherited in Gaussian process emulation results, which are called hereditary errors, and cannot be eliminated unless more accurate training data are used. This type of hereditary errors can be avoided only by using training data provided by a full finite element model.

Multiple fidelity data modeling has emerged from applications of Gaussian processes to realistic situations where simulation model calibration, verification, and validation should be taken into consideration (Joshi and Neat, 1998; Thacker et al., 2004; Kennedy

and O'Hagan, 2001). When data from sources of different fidelities are used for Gaussian process regression, the efficiency of predicting at a high level of accuracy can be greatly improved (Kennedy and O'Hagan, 2000a). A typical two-level case is that there are many low-fidelity data but only a few high-fidelity data are available. In this thesis a CMS model and a full finite element model are used to generate the low-fidelity and high-fidelity data, respectively. Two-level Gaussian processes are introduced for simultaneously minimizing hereditary errors and inference errors. A computationally cheap CMS model provides the low-fidelity data sufficient enough for minimizing inference errors in predictions given by the Gaussian process regression model at the low-fidelity level. The differences between data at different fidelity levels are assumed to be normally distributed and the difference between the distributions of low-fidelity predictions and of high-fidelity predictions appear to be a shift as a whole. This implies that the hereditary errors are the major errors in the predictions given by low-fidelity data. Such hereditary errors can be reduced by the additional information contained in the high-fidelity data. Thus, two-level Gaussian processes offer both the satisfactory accuracy and the high efficiency for characterizing structural dynamic responses with uncertainties.

The rest of the chapter is organized as follows: in Chapter 4.2 a thorough description of two-level Gaussian processes is presented based on the basics about Gaussian processes presented in Chapter 3; numerical examples are given in Chapter 4.3 followed by concluding remarks in Chapter 4.4.

4.2 Two-Level Gaussian Processes

4.2.1 Model of two-level Gaussian process regression

Gaussian processes are a well-established statistical tool for modeling various spatial and temporal problems. While there is a rich research literature on general theory (Mackay, 2003; Rasmussen and Williams, 2006), interested readers may refer to the paper by the author (Xia and Tang 2011) for applications of single-level Gaussian processes to structural dynamic analysis. As pointed out in Introduction, accuracy of single-level Gaussian processes is not only determined by inference method itself but also influenced by the fidelity of training data. A typical situation in uncertainty analysis is that we have only a few high-fidelity data and relatively many low-fidelity data, which calls for an emulating approach that fully uses information from the both datasets. Two-level Gaussian process regression was proposed by Kennedy and O'Hagan (2000a, 2000b, 2001) and has been applied to engineering problems (Huang et al., 2006; Qian et al., 2006; Jiang and Mahadevan, 2009). Since in the literature so far we haven't find an easy-to-follow theoretical derivation of two-level Gaussian processes, an elaboration is presented in this section for general readers with engineering background.

Let a system be denoted as a function $y = f(\mathbf{x})$, where $\mathbf{x} \in R^m$ stands for an m-dimensional input vector. In reality, the exact value returned by $f(\mathbf{x})$ is rarely known but, while being subject to errors, can be observed analytically, numerically, or experimentally. Let such an observed value, called output, be denoted as y $(y \in R)$. Given a set of n observations of the same fidelity, a dataset is defined as $\mathbf{D} = \{(y_i, \mathbf{x}_i) \mid i = 1, 2, ..., n\}$, in which each $(y_i, \mathbf{x}_i)$ is called a data point. Single-level Gaussian process regression can make prediction $y*$ at a target point $\mathbf{x}*$ given a data set **D**. Now, consider two datasets at different fidelity levels

$$\boldsymbol{D}^{(m)} = \left\{(\mathbf{x}_i^{(m)}, y_i^{(m)}) \mid i = 1, 2, ..., n^{(m)}\right\} \quad (m = 1, 2) \tag{4-1}$$

where n is the number of data points and the superscript m distinguishes between low-fidelity (m = 1) from high-fidelity (m = 2). The goal is to make prediction for $y^{(2)}*$ at a target point $\mathbf{x}^{(2)}*$ given a low-fidelity dataset $\boldsymbol{D}^{(1)}$ and a high-fidelity dataset $\boldsymbol{D}^{(2)}$. While

the basic idea behind single-level Gaussian processes is to extend the multivariate Gaussian distribution of a discrete subset of the range of a function to its entire continuous range, two-level Gaussian processes go further to link two functions at different fidelity levels using an assumed quasi-linear relation, which can be expresses as

$$y^{(1)} = \delta^{(1)} \tag{4-2}$$

$$y^{(2)} = \rho^{(1)} y^{(1)} + \delta^{(2)} \tag{4-3}$$

where $\rho^{(1)}$ is a regression parameter and $\delta^{(1)}$ and $\delta^{(2)}$ are independent stationary Gaussian processes

$$\delta^{(m)}(\mathbf{X}^{(m)}) \sim \mathbf{N}(\boldsymbol{\mu}^{(m)}(\mathbf{X}^{(m)}), \boldsymbol{\Sigma}^{(m)}(\mathbf{X}^{(m)}, \mathbf{X}^{(m)})) \tag{4-4}$$

where the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are, respectively, defined as

$$\boldsymbol{\mu}^{(m)} = \begin{bmatrix} \mu_1^{(m)} \\ \mu_2^{(m)} \\ \vdots \\ \mu_{n^{(m)}}^{(m)} \end{bmatrix}, \quad \boldsymbol{\Sigma}^{(m)} = \begin{bmatrix} k_{11}^{(m)} & \cdots & k_{1n^{(m)}}^{(m)} \\ \vdots & \ddots & \vdots \\ k_{n^{(m)}1}^{(m)} & \cdots & k_{n^{(m)}n^{(m)}}^{(m)} \end{bmatrix} \tag{4-5}$$

where the elements of the two matrices, known as the mean and covariance functions, respectively, are assumed to be (Kennedy and O'Hagan, 2000a)

$$\mu_i^{(m)} = h(\mathbf{x}_i^{(m)})\boldsymbol{\beta}^{(m)} \tag{4-6}$$

$$k_{ij}^{(m)} = \sigma_f^{(m)2} \exp\left\{-b^{(m)}(\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)})^T(\mathbf{x}_i^{(m)} - \mathbf{x}_j^{(m)})\right\} + \sigma_n^{(m)2}\delta_{ij}(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) \tag{4-7}$$

where $i, j = 1, 2, ..., n^{(m)}$, $k_{ij}^{(m)} = k_{ji}^{(m)}$, $h(\mathbf{x})$ is a base function vector to be determined in sequential discussions, $\boldsymbol{\beta}^{(m)}$ is a parameter vector with respect to the base function vector, parameters $\sigma_f^{(m)}, \sigma_n^{(m)}$, and $b^{(m)}$, which are associated with the covariance function, are termed as hyper-parameters, and δ_{ij} is the Kronecker delta. The input points of the high-fidelity dataset $\mathbf{X}^{(2)}$ are selected so as to satisfy the relation $\mathbf{X}^{(2)} \subseteq \mathbf{X}^{(1)}$ ($\mathbf{X}^{(m)} \triangleq \begin{bmatrix} \mathbf{x}_1^{(m)} & \mathbf{x}_2^{(m)} & \cdots & \mathbf{x}_{n^{(m)}}^{(m)} \end{bmatrix}^{\mathrm{T}}, m = 1, 2$). Therefore, the notation of covariance function $k_{ij}^{(1)}$, as defined in Eq(4-7) for m = 1, can be used for two data points either in the dataset of low-fidelity or in two datasets of different fidelities, namely both $k_{ij}^{(1)}(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(1)})$ and

$k_{ij}^{(1)}(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(2)})$ hold just because of $\mathbf{X}^{(2)} \subseteq \mathbf{X}^{(1)}$. A favorable property of Gaussian process regression is that the sum of independent Gaussians, e.g., $\delta^{(1)}$ and $\delta^{(2)}$, remains in a closed form such that the formulas for both the mean and covariance of high-fidelity prediction at target point can be expressed explicitly.

In order to work out the distribution of the high-fidelity predictions $y^{(2)}$ as defined in Eq (4-3), we first apply the affine transformation to the first term $\rho^{(1)} y^{(1)}$ on the right side of Eq (4-3)

$$\rho^{(1)}\delta^{(1)}(\mathbf{X}^{(2)}) \sim \mathbf{N}\left(\rho^{(1)} h^{\mathrm{T}}(\mathbf{X}^{(2)})\boldsymbol{\beta}^{(1)}, \rho^{(1)2}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)})\right) \tag{4-8}$$

where the legitimacy of the notation $\delta^{(1)}(\mathbf{X}^{(2)})$ is also due to the relation $\mathbf{X}^{(2)} \subseteq \mathbf{X}^{(1)}$. Then, following the rule for summing two independent random variables and substituting $\delta^{(2)}(\mathbf{X}^{(2)})$ defined in Eq (4-4) and $\rho^{(1)}\delta^{(1)}(\mathbf{X}^{(2)})$ in Eq (4-8) into Eq (4-3), it can be shown that

$$y^{(2)} \sim \mathbf{N}\left(\rho^{(1)} h^{T}(\mathbf{X}^{(2)})\boldsymbol{\beta}^{(1)} + h^{\mathrm{T}}(\mathbf{X}^{(2)})\boldsymbol{\beta}^{(2)}, \rho^{(1)2}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) + \boldsymbol{\Sigma}^{(2)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)})\right) \tag{4-9}$$

With Eq (4-2) and (4-9), it is convenient to represent both the low-fidelity data points in $\boldsymbol{D}^{(1)}$ and the high-fidelity data points in $\boldsymbol{D}^{(2)}$ in the form of a multivariate Gaussian distribution

$$\begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} \boldsymbol{\mu}^{(1)}(\mathbf{X}^{(1)}) \\ \boldsymbol{\mu}^{(2)}(\mathbf{X}^{(2)}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^{(1,1)}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) & \boldsymbol{\Sigma}^{(1,2)}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \\ \boldsymbol{\Sigma}^{(1,2)}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)})^{\mathbf{T}} & \boldsymbol{\Sigma}^{(2,2)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) \end{bmatrix}\right) \tag{4-10}$$

where

$$\mathbf{y}^{(m)} = [y_1^{(m)}, y_2^{(m)} ..., y_{n^{(m)}}^{(m)}]^{\mathbf{T}} \quad (m = 1, 2) \tag{4-11}$$

$$\boldsymbol{\mu}^{(m)}(\mathbf{X}^{(m)}) = \begin{bmatrix} h^T(\mathbf{x}_1^{(m)})\boldsymbol{\beta}^{(m)} \\ \vdots \\ h^T(\mathbf{x}_{n^{(m)}}^{(m)})\boldsymbol{\beta}^{(m)} \end{bmatrix} \tag{4-12}$$

$$\boldsymbol{\Sigma}^{(1,1)}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) = \boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(1)}, \mathbf{X}^{(1)}) \tag{4-13}$$

$$\boldsymbol{\Sigma}^{(2,2)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) = \rho^{(1)2}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) + \boldsymbol{\Sigma}^{(2)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}) \tag{4-14}$$

$$\boldsymbol{\Sigma}^{(1,2)}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) = \rho^{(1)}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}) \tag{4-15}$$

The expression of $\mathbf{\Sigma}^{(1,2)}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})$ in Eq (4-15) cannot be obtained as easily as that in Eq (4-13) and (4-14). To understand it, we need to go back to the original definition for covariance

$$Cov\left(y_i^{(1)},y_j^{(2)}\right)=E[y_i^{(1)}y_j^{(2)}]-E[y_i^{(1)}]E[y_j^{(2)}] \tag{4-16}$$

where $E[\cdot]$ denotes expected value or mean. From the definitions in Eqs (4-4) to (4-6) we notice that $E[\delta^{(m)}(\mathbf{x}_i^{(m)})]\equiv h^T(\mathbf{x}_i^{(m)})\boldsymbol{\beta}^{(m)}$. Thus, $E[y_i^{(1)}]$ and $E[y_j^{(2)}]$ can be readily defined according to Eq (4-4) and (4-9), respectively

$$E[y_i^{(1)}]=h^T(\mathbf{x}_i^{(1)})\boldsymbol{\beta}^{(1)}=E[\delta^{(1)}(\mathbf{x}_i^{(1)})] \tag{4-17a}$$

$$E[y_j^{(2)}]=\rho^{(1)}h^T(\mathbf{x}_j^{(2)})\boldsymbol{\beta}^{(1)}+h^T(\mathbf{x}_j^{(2)})\boldsymbol{\beta}^{(2)}=\rho^{(1)}E[\delta^{(1)}(\mathbf{x}_j^{(2)})]+E[\delta^{(2)}(\mathbf{x}_j^{(2)})] \tag{4-17b}$$

where, again, $\delta^{(1)}(\mathbf{x}_j^{(2)})$ holds just because of $\mathbf{X}^{(2)}\subseteq\mathbf{X}^{(1)}$. To find an appropriate expression for $E[y_i^{(1)}y_j^{(2)}]$ in Eq (4-16), let us explicitly write down the multiplication of the two independent random variables $y_i^{(1)}$ and $y_j^{(2)}$

$$y^{(1)}y^{(2)}=\rho^{(1)}\delta^{(1)}\delta^{(1)}+\delta^{(1)}\delta^{(2)} \tag{4-18}$$

and its expected value

$$E[y_i^{(1)}y_j^{(2)}]=\rho^{(1)}E[\delta^{(1)}(\mathbf{x}_i^{(1)})\delta^{(1)}(\mathbf{x}_j^{(2)})]+E[\delta^{(1)}(\mathbf{x}_i^{(1)})]E[\delta^{(2)}(\mathbf{x}_j^{(2)})] \tag{4-19}$$

Substituting Eq (4-17) and (4-19) into Eq (4-16) gives rise to

$$Cov\left(y_i^{(1)},y_j^{(2)}\right)=\rho^{(1)}\left(E[\delta^{(1)}(\mathbf{x}_i^{(1)})\delta^{(1)}(\mathbf{x}_j^{(2)})]-E[\delta^{(1)}(\mathbf{x}_i^{(1)})]\cdot E[\delta^{(1)}(\mathbf{x}_j^{(2)})]\right) \tag{4-20}$$

Look at the expression inside the parenthesis on the right side: it is exactly the original definition of covariance between $\delta^{(1)}(\mathbf{x}_i^{(1)})$ and $\delta^{(1)}(\mathbf{x}_j^{(2)})$. So, we have

$$Cov\left(y_i^{(1)},y_j^{(2)}\right)=\rho^{(1)}Cov\left(\delta^{(1)}(\mathbf{x}_i^{(1)}),\delta^{(1)}(\mathbf{x}_j^{(2)})\right)=\rho^{(1)}Cov\left(y^{(1)}(\mathbf{x}_i^{(1)}),y^{(1)}(\mathbf{x}_j^{(2)})\right)=\rho^{(1)}k^{(1)}(\mathbf{x}_i^{(1)},\mathbf{x}_j^{(2)}) \tag{4-21}$$

where the definition in Eq (4-2) has been used and, again, $y^{(1)}(\mathbf{x}_j^{(2)})$ holds because of $\mathbf{X}^{(2)}\subseteq\mathbf{X}^{(1)}$. Thus, the explicit expression of $\mathbf{\Sigma}^{(1,2)}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})$ is found

$$\mathbf{\Sigma}^{(1,2)}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})=\rho^{(1)}\mathbf{\Sigma}^{(1)}(\mathbf{X}^{(1)},\mathbf{X}^{(2)}) \tag{4-15}$$

where the notation $\mathbf{\Sigma}^{(1)}(\mathbf{X}^{(1)},\mathbf{X}^{(2)})$ is legitimate because, as explained previously, $k_{ij}^{(1)}(\mathbf{x}_i^{(1)},\mathbf{x}_j^{(2)})$ holds because of $\mathbf{X}^{(2)}\subseteq\mathbf{X}^{(1)}$.

Since Bayesian approach is used for parameter estimation, we now re-write Eq (4-10) in an explicitly Bayesian fashion

$$\mathbf{y} \mid \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \boldsymbol{\beta}, \phi \sim \mathbf{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4-22}$$

where a set of compact notations are defined as

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \end{bmatrix}, \ \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}^{(1)}(\mathbf{X}^{(1)}) \\ \boldsymbol{\mu}^{(2)}(\mathbf{X}^{(2)}) \end{bmatrix}, \ \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}^{(1,1)} & \boldsymbol{\Sigma}^{(1,2)} \\ \boldsymbol{\Sigma}^{(1,2)\mathbf{T}} & \boldsymbol{\Sigma}^{(2,2)} \end{bmatrix} \tag{4-23a, b, c}$$

Notice that Eq (4-10) and Eq (4-22) only involve the data points in both $\boldsymbol{D}^{(1)}$ and $\boldsymbol{D}^{(2)}$. Predictions of high-fidelity outputs $\mathbf{y}^{(2)}*$ at target points $\mathbf{X}^{(2)}*$, which is a subset of $\mathbf{X}* = \mathbf{X}^{(1)} - \mathbf{X}^{(2)}$, also need to be included in the standard form of Gaussian processes (Rasmussen and Williams, 2006)

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^{(2)}* \end{bmatrix} \sim \mathbf{N}\left(\begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu}^{(2)}* \end{bmatrix} \ , \ \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\Sigma}* \\ \boldsymbol{\Sigma}*^{\mathbf{T}} & \boldsymbol{\Sigma}^{(2,2)}** \end{bmatrix} \right) \tag{4-24}$$

where

$$\boldsymbol{\Sigma}* = \begin{bmatrix} \rho^{(1)}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(1)}, \mathbf{X}^{(2)}*) \\ \rho^{(1)2}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}*) + \boldsymbol{\Sigma}^{(2)}(\mathbf{X}^{(2)}, \mathbf{X}^{(2)}*) \end{bmatrix} \tag{4-25}$$

$$\boldsymbol{\Sigma}**^{(2,2)} = \rho^{(1)2}\boldsymbol{\Sigma}^{(1)}(\mathbf{X}^{(2)}*, \mathbf{X}^{(2)}*) + \boldsymbol{\Sigma}^{(2)}(\mathbf{X}^{(2)}*, \mathbf{X}^{(2)}*) \tag{4-26}$$

The high-fidelity predictions $\mathbf{y}^{(2)}*$ at target points $\mathbf{X}*$ conditional on data points, parameters and hyper-parameters can be obtained

$$\mathbf{y}^{(2)}* \mid \mathbf{X}*, \mathbf{X}^{(1)}, \mathbf{y}^{(1)}, \mathbf{X}^{(2)}, \mathbf{y}^{(2)}, \boldsymbol{\beta}, \phi \sim \mathbf{N}\left(\boldsymbol{\mu}^{(2)}* + \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}), \boldsymbol{\Sigma}**^{(2,2)} - \boldsymbol{\Sigma}*^{\mathbf{T}} \boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}* \right) \tag{4-27}$$

where the set of all parameters and the set of all hyper-parameters are denoted, respectively, as $\boldsymbol{\beta} = [\boldsymbol{\beta}^{(1)} \quad \boldsymbol{\beta}^{(2)}]^{\mathrm{T}}$ and $\phi = \{\rho^{(1)}, \sigma_f^{(1)}, b^{(1)}, \sigma_n^{(1)}, \sigma_f^{(2)}, b^{(2)}, \sigma_n^{(2)}\}$.

4.2.2 Bayesian inference for $\boldsymbol{\beta}$

Within the Bayesian framework we should update both the parameters $\boldsymbol{\beta}$ and hyper-parameters ϕ in Eq (4-27) with the information contained in the data $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}\}$ so as to obtain posterior predictions at target points $\mathbf{y}^{(2)}* \mid \mathbf{X}*, \mathbf{X}^{(1)}, \mathbf{y}^{(1)}, \mathbf{X}^{(2)}, \mathbf{y}^{(2)}, \hat{\boldsymbol{\beta}}, \hat{\phi}$, in which the hat notation stands for 'posterior'. Because of the difficulty of evaluating the integrals in

a full Bayesian approach, a strategy of hierarchical modeling is adopted to estimate parameters and hyper-parameters separately at two stages. Ideally, the posterior hyper-parameters $\hat{\phi}$ are estimated only conditional on the data $\{\mathbf{D}^{(1)},\mathbf{D}^{(2)}\}$ at the first stage and then at the second stage the posterior parameters $\hat{\boldsymbol{\beta}}$ are determined conditional on both $\{\mathbf{D}^{(1)},\mathbf{D}^{(2)}\}$ and $\hat{\phi}$. In practice, such a strategy allows the flexibility of evaluating hyper-parameters by maximum likelihood estimation instead of numerically more difficult Bayesian inference (Kennedy and O'Hagan, 2000b; Xia and Tang, 2011).

As Bayesian parameter estimation required, we should in the first place specify distributions to represent our prior knowledge. However, the fact is that little information is known in advance regarding parameters $\boldsymbol{\beta}$ and hyper-parameters ϕ. Following the common practice in this typical situation (Jaynes, 2003), we assign to them improper priors and independent non-informative priors, respectively (Kennedy and O'Hagan, 2000a)

$$P(\beta^{(1)},\sigma_f^{(1)2},b^{(1)},\sigma_n^{(1)2}) \propto \sigma_f^{(1)-2}b^{(1)-1}\sigma_n^{(1)-2} \tag{4-28}$$

$$P(\beta^{(2)},\rho^{(1)},\sigma_f^{(2)2},b^{(2)},\sigma_n^{(2)2}) \propto \rho^{(2)-1}\sigma_f^{(2)-2}b^{(2)-1}\sigma_n^{(2)-2} \tag{4-29}$$

Then the posterior distribution of $\boldsymbol{\beta}$ and ϕ given the data $\{\mathbf{D}^{(1)},\mathbf{D}^{(2)}\}$ can be obtained

$$P(\boldsymbol{\beta},\phi \mid \mathbf{y}^{(1)},\mathbf{X}^{(1)},\mathbf{y}^{(2)},\mathbf{X}^{(2)}) = \frac{P(\mathbf{y} \mid \mathbf{X}^{(1)},\mathbf{X}^{(2)},\boldsymbol{\beta},\phi)P(\boldsymbol{\beta},\phi)}{P(\mathbf{y} \mid \mathbf{X}^{(1)},\mathbf{X}^{(2)})} \propto P(\mathbf{y} \mid \mathbf{X}^{(1)},\mathbf{X}^{(2)},\boldsymbol{\beta},\phi)$$

$$\propto \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\} \tag{4-30}$$

where $P(\mathbf{y} \mid \mathbf{X}^{(1)},\mathbf{X}^{(2)},\boldsymbol{\beta},\phi)$ is given in Eq (4-22) and $(\mathbf{y}-\boldsymbol{\mu})^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})$ can be expanded with the notations defined in Eq (4-23) and re-written in a quadratic form with respect to $\boldsymbol{\beta}$

$$(\mathbf{y}-\boldsymbol{\mu})^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}) = \left(\boldsymbol{\beta}-(\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}\right)^{\mathbf{T}}\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H}\left(\boldsymbol{\beta}-(\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}\right) \tag{4-31}$$

where

$$\mathbf{H}=\begin{bmatrix} h^T(\mathbf{x}_1^{(1)}) & \mathbf{0} \\ \vdots & \vdots \\ h^T(\mathbf{x}_{n1}^{(1)}) & \mathbf{0} \\ \rho^{(1)}h^T(\mathbf{x}_1^{(2)}) & h^T(\mathbf{x}_1^{(2)}) \\ \vdots & \vdots \\ \rho^{(1)}h^T(\mathbf{x}_{n2}^{(2)}) & h^T(\mathbf{x}_{n2}^{(2)}) \end{bmatrix} \tag{4-32}$$

Note that, as far as $\boldsymbol{\beta}$ is concerned at the second stage where ϕ is assumed to be known within the two-stage strategy, Eq (4-31) , which now has puts on a standard exponential form of Gaussian distribution, can be exactly interpreted as the distribution of parameters $\boldsymbol{\beta}$ conditional on the hyper-parameters ϕ and the data $\{\mathbf{D}^{(1)},\mathbf{D}^{(2)}\}$ (Lindley and Smith, 1972)

$$\boldsymbol{\beta}\,|\,\mathbf{y}^{(1)},\mathbf{X}^{(1)},\mathbf{y}^{(2)},\mathbf{X}^{(2)},\phi \sim \mathbf{N}\left(\hat{\boldsymbol{\beta}},\mathbf{W}\right) \tag{4-33}$$

where the posterior parameters $\hat{\boldsymbol{\beta}}=(\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{y}$ and covariance matrix $\mathbf{W}=(\mathbf{H}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}$. It is the properties of both Gaussian and improper prior that lead to this closed-form expression of the distribution of $\boldsymbol{\beta}$. However, it is more complicated to update the hyper-parameters ϕ with the data $\{\mathbf{D}^{(1)},\mathbf{D}^{(2)}\}$, which will be discussed in the immediately following part.

4.2.3 Posterior predictions of high fidelity

The flexibility of the two-stage parameter estimation strategy allows the hyper-parameters ϕ to be evaluated by maximum likelihood principle. In contrast with a full Bayesian approach where usually numerical integrals have to be done by numerically costly Markov chain Monte Carlo (MCMC) methods (Andrieu et al., 2003), maximum likelihood estimation is relatively easy to be implemented. Thus, the hierarchical is an empirical semi-Bayesian approach, and estimation of ϕ will be described in the next section while in this section we, assuming $\hat{\phi}$ is already known as well as $\hat{\boldsymbol{\beta}}$, continue to finish the derivation of the posterior predictions of high fidelity within the Bayesian

framework. It begins with a lemma which proof can be found in the paper by Lindley and Smith (1972).

Lemma. Suppose that, given a set of parameters $\boldsymbol{\theta}_1$, a Gaussian is defined

$$\mathbf{y} \sim \mathbf{N}\left(\mathbf{A}_1\boldsymbol{\theta}_1, \mathbf{C}_1\right) \tag{4-34}$$

and, given another set of parameters $\boldsymbol{\theta}_2$, another Gaussian is also defined

$$\boldsymbol{\theta}_1 \sim \mathbf{N}\left(\mathbf{A}_2\boldsymbol{\theta}_2, \mathbf{C}_2\right) \tag{4-35}$$

then the marginal distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim \mathbf{N}\left(\mathbf{A}_1\mathbf{A}_2\boldsymbol{\theta}_2, \mathbf{C}_1 + \mathbf{A}_1\mathbf{C}_2\mathbf{A}_1^{\mathbf{T}}\right) \tag{4-36}$$

where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{C}_1$ and $\mathbf{C}_2$ are known positive-definite matrices of correspondingly proper dimensions.

The purpose of introducing this lemma is to integrate out the first set of parameters $\boldsymbol{\theta}_1$ in evaluating the posterior $\mathbf{y}$ as seen in Eq (4-36). Following this lemma, the first step is to list the two equations obtained above as given conditions corresponding to Eq (4-34) and (4-35), respectively

$$\mathbf{y}^{(2)}* \mid \mathbf{X}*, \mathbf{X}^{(1)}, \mathbf{y}^{(1)}, \mathbf{X}^{(2)}, \mathbf{y}^{(2)}, \boldsymbol{\beta}, \phi \sim \mathbf{N}\left(\boldsymbol{\mu}^{(2)}* + \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu}), \boldsymbol{\Sigma}**^{(2,2)} - \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}*\right) \tag{4-27}$$

$$\boldsymbol{\beta} \mid \mathbf{y}^{(1)}, \mathbf{X}^{(1)}, \mathbf{y}^{(2)}, \mathbf{X}^{(2)}, \phi \sim \mathbf{N}\left(\hat{\boldsymbol{\beta}}, \mathbf{W}\right) \tag{4-33}$$

It is not difficult to verify that the mean term in Eq (4-27) can be expanded and then re-expressed as a product that explicitly contains $\boldsymbol{\beta}$ in the same way as $\boldsymbol{\theta}_1$ in Eq (4-34). Therefore, Eq (4-27) and (4-33) obviously meet the required conditions stated in Eq (4-34) and (4-35), which simply gives rise to

$$\mathbf{y}^{(2)}* \mid \mathbf{X}*, \mathbf{X}^{(1)}, \mathbf{y}^{(1)}, \mathbf{X}^{(2)}, \mathbf{y}^{(2)}, \phi$$

$$\sim \mathbf{N}\left(\mathbf{H}'\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}}), \boldsymbol{\Sigma}**^{(2,2)} - \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}* + (\mathbf{H}' - \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H})\mathbf{W}(\mathbf{H}'(\mathbf{X}*) - \boldsymbol{\Sigma}*^{\mathbf{T}}\,\boldsymbol{\Sigma}^{-1}\mathbf{H})^{\mathbf{T}}\right) \tag{4-37}$$

where

$$\mathbf{H}' = \begin{bmatrix} \rho^{(1)}h^T(\mathbf{x}_1*) & h^T(\mathbf{x}_1*) \\ \vdots & \vdots \\ \rho^{(1)}h^T(\mathbf{x}_{n*}*) & h^T(\mathbf{x}_{n*}*) \end{bmatrix} \tag{4-38}$$

and

$$\mathbf{H}=\begin{bmatrix} h^T(\mathbf{x}_1^{(1)}) & \mathbf{0} \\ \vdots & \vdots \\ h^T(\mathbf{x}_{n^{(2)}}^{(1)}) & \mathbf{0} \\ \rho^{(1)}h^T(\mathbf{x}_1^{(2)}) & h^T(\mathbf{x}_1^{(2)}) \\ \vdots & \vdots \\ \rho^{(1)}h^T(\mathbf{x}_{n^{(2)}}^{(2)}) & h^T(\mathbf{x}_{n^{(2)}}^{(2)}) \end{bmatrix} \tag{4-39}$$

Note that Eq (4-37), in which parameters $\boldsymbol{\beta}$ has been integrated out, is to be used for calculating means and credible intervals of the posterior high-fidelity predictions $\mathbf{y}^{(2)}*$ at target points $\mathbf{X}*$ as long as hyper-parameters are determined ϕ.

4.2.4 Maximum likelihood estimation of ϕ

Because of the mentioned numerical difficulty of integrals required by a full Bayesian approach, hierarchical modeling has been adopted for evaluating parameters and hyper-parameters separately. It should be clear that such a two-hierarchy framework refers to estimating parameter and hyper-parameters hierarchically separately, but not inferencing with data at two fidelity levels. Indeed, even for single-level Gaussian process regression (Xia and Tang, 2011) parameters and hyper-parameters can be estimated using the two-stage strategy. Now, for two-level Gaussian process regression, parameters and hyper-parameters also can be estimated separately at two different stages corresponding to the two levels of data fidelity. That is procedures: the work at the first stage is to derive the posterior parameters $\hat{\boldsymbol{\beta}}$, which has been described in Section 3.2, and furthermore, corresponding to the two levels of data fidelity, at the second stage two separate steps are taken to estimate hyper-parameters $\phi=\{\phi_1,\phi_2\}$, where first subset of hyper-parameters $\phi_1=\{\sigma_f^{(1)},b^{(1)},\sigma_n^{(1)}\}$ are estimated with only the low-fidelity data, and then the second subset of hyper-parameters $\phi_2=\{\rho^{(1)},\sigma_f^{(2)},b^{(2)},\sigma_n^{(2)}\}$ are estimated with the estimated $\hat{\phi}_1$ and data at both the two fidelity levels (Kennedy and O'Hagan, 2000b). Following Xia and Tang (2011), estimation of ϕ_1 is determined by maximizing the logarithm marginal likelihood (Kennedy and O'Hagan, 2000a)

$$L = -\frac{n^{(1)}}{2}\log 2\pi - \frac{1}{2}\log|\boldsymbol{\Sigma}^{(1,1)}| - \frac{1}{2}[\mathbf{y}^{(1)} - \mathbf{h}(\mathbf{X}^{(1)})\tilde{\boldsymbol{\beta}}^{(1)}]^{\mathrm{T}}\boldsymbol{\Sigma}^{(1,1)-1}[\mathbf{y}^{(1)} - \mathbf{h}(\mathbf{X}^{(1)})\tilde{\boldsymbol{\beta}}^{(1)}] \quad (4\text{-}40)$$

where $\mathbf{h}(\mathbf{X}^{(1)}) = \left[h(\mathbf{x}_1^{(1)}) \quad h(\mathbf{x}_2^{(1)}) \quad \cdots \quad h(\mathbf{x}_{n^{(1)}}^{(1)}) \right]^{\mathrm{T}}$ and $\tilde{\boldsymbol{\beta}}^{(1)}$ is the posterior parameters with respect to the low-fidelity level

$$\tilde{\boldsymbol{\beta}}^{(1)} = (\mathbf{h}(\mathbf{X}^{(1)})^{\mathrm{T}}\boldsymbol{\Sigma}^{(1,1)-1}\mathbf{h}(\mathbf{X}^{(1)}))^{-1}\mathbf{h}(\mathbf{X}^{(1)})^{\mathrm{T}}\boldsymbol{\Sigma}^{(1,1)-1}\mathbf{y}^{(1)} \quad (4\text{-}41)$$

At the second stage $\hat{\phi}_1$ and $\tilde{\boldsymbol{\beta}}^{(1)}$ are known as given values. Re-write Eq (4-3) in an explicit form

$$\mathbf{d} = \begin{bmatrix} y_1^{(2)} - \rho^{(1)}\mathrm{h}(x_1^{(2)})\tilde{\boldsymbol{\beta}}^{(1)} \\ \vdots \\ y_{n^{(2)}}^{(2)} - \rho^{(1)}\mathrm{h}(x_{n^{(2)}}^{(2)})\tilde{\boldsymbol{\beta}}^{(1)} \end{bmatrix} \quad (4\text{-}42)$$

which actually has been defined in Eq (4-4) as a Gaussian such that the hyper-parameters ϕ_2 at the high-fidelity level can also be determined by maximizing the logarithm marginal likelihood (Kennedy and O'Hagan, 2000a)

$$-\frac{n^{(2)}}{2}\log 2\pi - \frac{1}{2}\log\left|\boldsymbol{\Sigma}^{(2,2)}\right| - \frac{1}{2}(\mathbf{d} - \mathbf{h}(\mathbf{X}^{(2)})\tilde{\boldsymbol{\beta}}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}^{(2,2)-1}(\mathbf{d} - \mathbf{h}(\mathbf{X}^{(2)})\tilde{\boldsymbol{\beta}}^{(2)}) \quad (4\text{-}43)$$

where $\mathbf{h}(\mathbf{X}^{(2)}) = \left[h(\mathbf{x}_1^{(2)}) \quad h(\mathbf{x}_2^{(2)}) \quad \cdots \quad h(\mathbf{x}_{n^{(2)}}^{(2)}) \right]^{\mathrm{T}}$ and $\tilde{\boldsymbol{\beta}}^{(2)}$ is the posterior parameters with respect to the high-fidelity level

$$\tilde{\boldsymbol{\beta}}^{(2)} = (\mathbf{h}(\mathbf{X}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}^{(2,2)-1}\mathbf{h}(\mathbf{X}^{(2)}))^{-1}\mathbf{h}(\mathbf{X}^{(2)})^{\mathrm{T}}\boldsymbol{\Sigma}^{(2,2)-1}\mathbf{d} \quad (4\text{-}44)$$

Here, for simplicity, we still use the hat notation to denote the hyper-parameters $\hat{\phi}_1$ and $\hat{\phi}_2$ that are determined in Eq (4-40) and (4-43) by maximum likelihood estimation though it generally stands for 'posterior'. The posterior parameters $\hat{\boldsymbol{\beta}}$ can be evaluated simply by substituting $\hat{\phi} = \{\hat{\phi}_1, \hat{\phi}_2\}$ together with data $\{\mathbf{D}^{(1)}, \mathbf{D}^{(2)}\}$ into Eq (4-33). Finally, high-fidelity predictions $\mathbf{y}^{(2)}*$ at target points $\mathbf{X}*$ are calculated using Eq (4-37) with $\hat{\phi}$ and $\hat{\boldsymbol{\beta}}$.

4.3 Numerical examples

There are three parts in this section: first, a full finite element model and a CMS model are built for an exemplary 3-component plate structure to serve as, respectively, a high-fidelity model and a low-fidelity model; second, the implementation of two-level Gaussian processes is illustrated by an example of emulating Monte Carlo simulation, which, from a general sense, is just the predicting of outputs at target points over an input space; finally, our discussions go back to the initial purpose of Monte Carlo simulation, showing that how two-level Gaussian process emulation helps to make more accurate statistical inferences.

4.3.1 Modeling at low- and high-fidelity levels

Let us consider a plate structure of which geometry and meshing for finite element modeling with the thin plate theory are displayed in Fig. 4-1. It is supposed to be made of homogenous and isotropic steel, which nominal mass density, Young's modulus, and Poisson ratio are, 7870kg/m^3, $2.10\times10^{11}\text{N/m}^2$, and 0.3, respectively. All boundaries of the full structure are free except that the left boundary is fixed. The full structure is divided into three substructures with free interface assumption. Substructure 1 has constraints at its left boundary and Substructure 2 and 3 are floating components having rigid body motion modes. Uncertainties are introduced in a parametric fashion: Young's modulus varies in each substructure and different substructures have different standard deviations of Young's modulus.

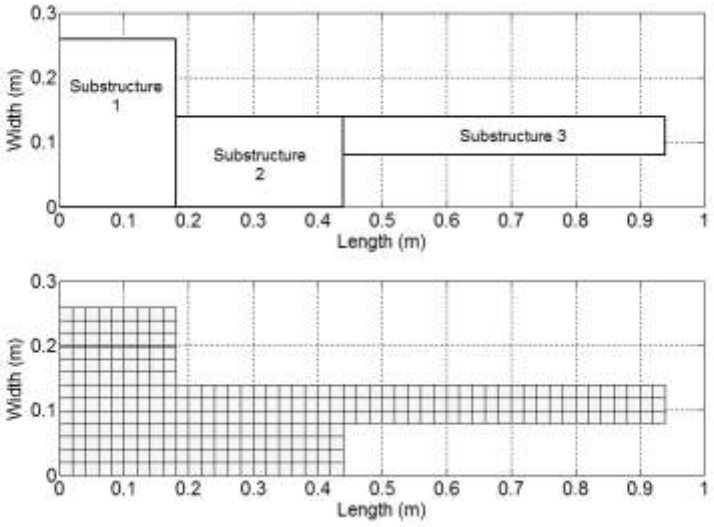

Fig. 4-1 Geometry and meshing of the 3-component plate structure

The high-fidelity finite element model is constructed on the structure as a whole while the low-fidelity model is constructed using free interface CMS based on finite element analyses on individual substructures. For simplicity, in this paper we refer to them as the full model and the CMS model, respectively. Accordingly, their results are taken as high-fidelity data and low-fidelity data, respectively, for two-level Gaussian process regression. The computational cost of the CMS model generally is much lower than the full model in terms of CPU times and memories. In fact, the low-fidelity data provided by a CMS model can represent those low-cost and error-prone data that are often much more available in practice than the expensive and accurate data that can also represented by the high-fidelity data provided by the full model.

A nominal model is defined as a model that put on the mean Young's modulus values of the substructures. The accuracy of the nominal CMS model can be inspected against the nominal full finite element model by comparing eigen-frequencies in Table 4-1. To

choose how many normal modes of substructures to be retained in CMS is a heuristic issue in practice. As a thumb of rule, however, it is enough to keep the normal modes of substructures up to at least 1.5 times the maximum frequency of the full structure that we are interested in. For illustration purposes, we keep 2, 6, and 6 normal modes of substructure 1, 2, and 3, respectively, in the CMS model, which means that the order of the reduced model is only 14.

Table 4-1 Comparisons of eigen-frequencies (Hz): full FE model versus CMS

Mode	Full FE model	CMS model	Relative Error (%)
1	3.01	2.96	1.37
2	12.95	12.92	0.19
3	33.12	32.77	1.08
4	53.61	54.71	2.06
5	63.53	63.87	0.53
6	96.33	103.94	7.90
7	109.15	107.12	1.86
8	126.44	138.30	9.39
9	163.30	170.25	4.25
10	214.80	231.14	7.60
11	231.90	249.69	7.67
12	296.93	350.59	18.07
13	306.95	438.58	42.88
14	340.93	3331.12	877.05

To see how the accuracy of model is affected by order reduction, let us compare the results of the CMS model with the full finite element model. First of all, it can be seen in Table 4-1 that the relative errors of the first five eigen-frequencies of the CMS model are less than 5%, which is usually acceptable in practice. Since only a few of substructure normal modes are used in the CMS model, the relative errors of eigen-frequencies of the CMS model soon begins to be conspicuous from the sixth mode. For instance, the relative

error of the sixth eigen-frequency of the CMS model has relative is as high as 7.9%. As seen in Table 4-1, the value of the sixth eigen-frequency of the CMS model is closer to the value of the seventh than the sixth eigen-frequency of the full finite element model. A question thus may arise why the sixth eigen-frequency of the CMS model is associated with the sixth, but not the seventh, eigen-frequency of the full finite element model. This is because the association of two modes is usually measured by the correlation between their mode shapes. In simple cases such as this example, the correlations of different pairs of modes can be compared by visual inspection of mode shapes. As plotted in Fig. 4-2, it is obvious that the mode shape of the sixth mode of the CMS model is more similar to that of the sixth than that of the seventh mode shapes of the full finite element model. A relatively more vigorous measure is to calculate and compare the values of modal assurance criterion (MAC).

Table 4-2 Comparisons of von Mises strains: full FE model versus CMS model

Mode	Full FE model	CMS model	Relative Error (%)
1	0.55	0.54	2.73
2	2.05	2.10	2.50
3	2.41	2.30	4.76
4	7.93	8.47	6.76
5	1.56	1.21	22.36
6	8.55	9.20	7.61
7	5.70	6.58	15.40
8	8.26	5.68	31.24
9	8.33	9.37	12.47
10	5.57	5.32	4.47
11	8.86	14.70	65.93
12	8.14	11.72	44.02
13	7.59	8.51	12.01
14	16.76	3.44	79.48

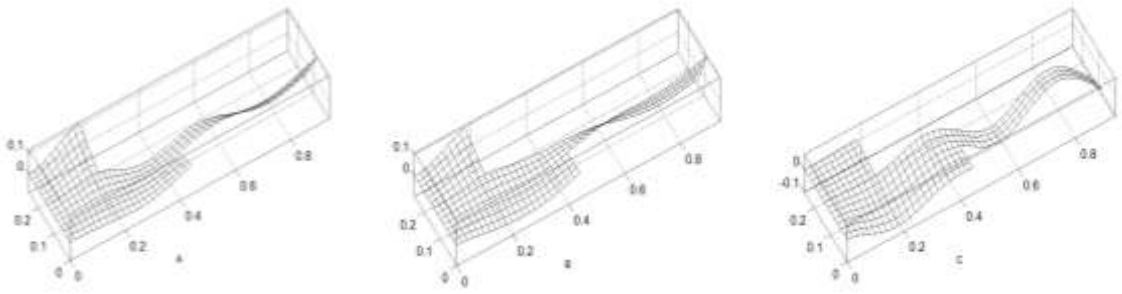

Fig. 4-2 Mode shape comparison: CMS versus Full FE. (a) the sixth mode of CMS model; (b) the sixth mode of full FE model; (c) the seventh mode of full FE model

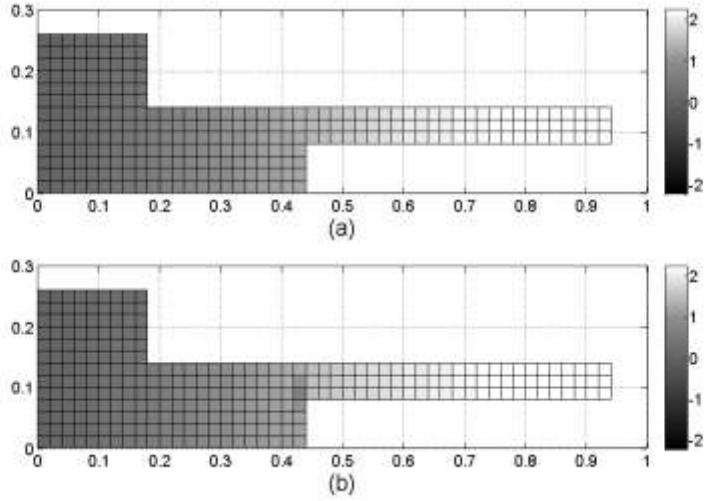

Fig. 4-3 Distribution of von Mises strains upon the first mode shape: (a) – full FE model; (b) – CMS model. Vertical axis: width (m); Horizontal axis: length (m)

In addition to eigen-frequencies, quantities closely related to mode shapes such as strains are also of great interest in design and analysis. Table 4-2 lists von Mises or equivalent strains of an element that is on the upper left corner of Substructure 2 with respect to different mode shapes. By comparing relative errors in Table 4-2 with Table 1,

it is evident that mode shapes related values are more sensitive to parameter variations than eigen-frequencies. For comparison purposes, Fig. 4-3 displays the distributions of von Mises strains with respect to the first mode shape of the full finite element model as well as the CMS model. In the following parts, both eigen-frequency and von Mises strain are taken as two representative quantities of interest to illustrate the applications of two-level Gaussian processes to all aspects of modal analysis.

4.3.2 Two-level Gaussian process emulation

Suppose Young's moduli in individual substructures, i.e., $E_1, E_2,$ and E_3 , are independent Gaussians having different standard deviations 2%, 5%, and 10%, respectively. Every particular configuration $\mathbf{E}=[E_1,E_2,E_3]^T$ is a multivariate Gaussian. Monte Carlo sampling is performed over the parameter configuration space, which is constitute of all possible **E,** using the full finite element model of structure as well as the CMS model. The results of the former are referred to as the high-fidelity data and that of the latter as the low-fidelity data. At each level of fidelity, a set of 1000 sample points is gathered by Latin hypercube design which ensures sample points spread evenly over the parameter configuration space and can be readily implemented using MATLAB®. The two sets of data are to be used as 'exact' values for assessing predictions obtained by Gaussian processes. From the 1000 sample points of the CMS model, 21 are selected as low-fidelity data points for the low-level Gaussian process regression by removing maximum ends among minimum distances of paired parameter configuration points. Detailed steps of this selecting algorithm that guarantees the selected data points be well-

spaced are described by Kennedy and O'Hagan (2000). In the same way, 3 high-fidelity data points are selected from the 1000 sample points of the full finite element model for the high-level Gaussian process regression.

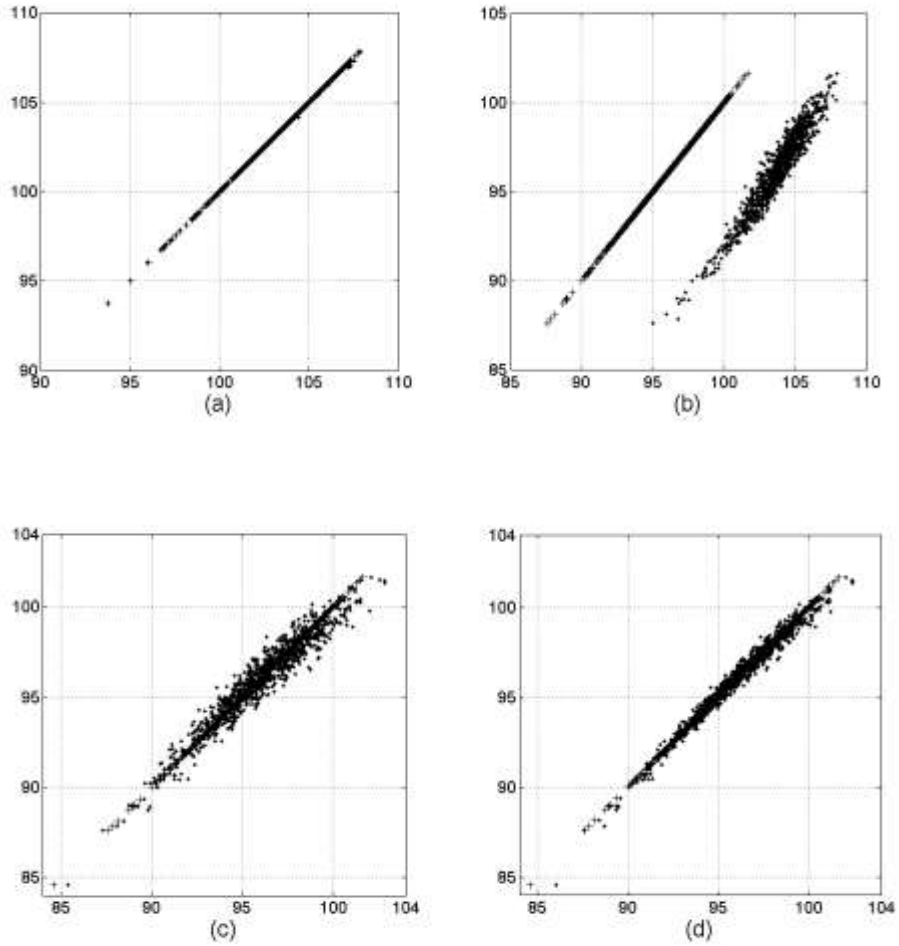

Fig. 4-4 (a) - Predictions using low-fidelity data points against low-fidelity sample points; (b) - predictions using low-fidelity data points against high-fidelity sample points; (c) - predictions using 21 low-fidelity data points against high-fidelity sample points; (d) predictions using 3 high-fidelity data points against high-fidelity sample points. Vertical axis: High-fidelity frequency (Hz); Horizontal axis: low-fidelity frequency (Hz). Cross: high-fidelity sample point; Dot: prediction

Fig. 4-4(a) and 6(a) display the respective frequency and von Mises strain predictions obtained using the 21 low-fidelity data points in the single-level Gaussian process regression model that has been introduced by the author in another paper (Xia and Tang, 2011). When compared against 'exact' values of low-fidelity, the two sets of predictions appear to be nearly perfect, which can be seen from the close matches in Fig. 4-4(a) and 6(a). If there is no concern about fidelity of data points and related predictions, the two examples sufficiently evidence the inference power of Gaussian processes. The reality, however, is that verification/validation of a simulation by comparing with data of higher-fidelity, whenever necessary, has to be carried out and, accordingly, some data of high-fidelity, though maybe very few, have to be acquired. Errors in results of a low-fidelity model may lead to inaccurate distribution of a sample either directly generated by a Monte Carlo simulation that employs such a low fidelity model or obtained by a Gaussian process emulator that uses data points provided by such a low-fidelity model. Thus, as compared with 'exact' values of the higher fidelity level, the predictions of a Gaussian process regression model that uses only low-fidelity data suffer become unreliable. This can be seen in Fig. 4-4(b) and 6(b), where distinct discrepancies exist between the 'exact' values of high-fidelity and the predictions based on low-fidelity data points. Fig. 4-5(a) presents the distribution of the two sets of values from another perspective, where an obvious systematic bias in the predictions based on low-fidelity data points can be observed. The primary source of such a bias in distribution is the approximations adopted in CMS modeling.

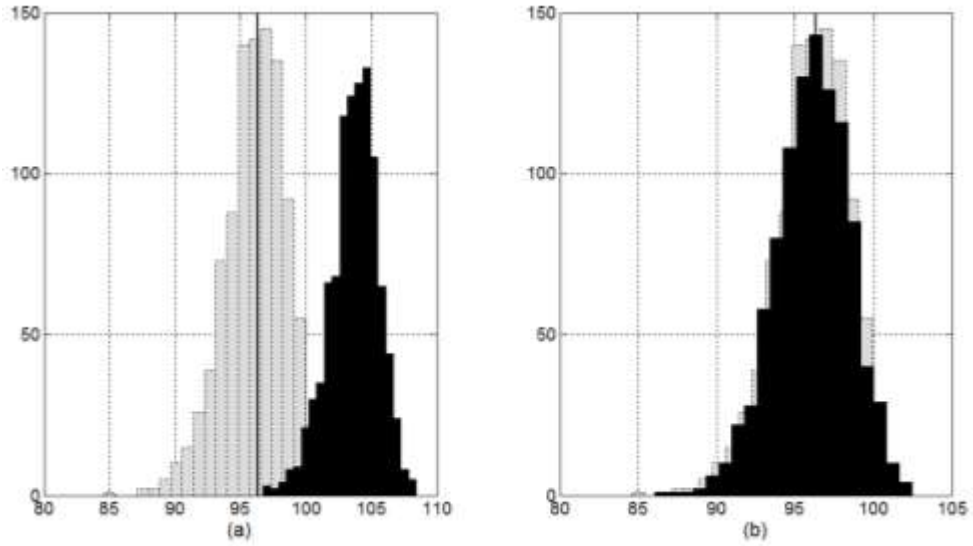

Fig. 4-5 Histogram of high-fidelity sample points versus (a) - predictions using 21 low-fidelity data points, and (b) - predictions using both 21 low-fidelity and 3 high-fidelity data points. Vertical axis: count; Horizontal axis: low-fidelity frequency (Hz). Grey patch: high-fidelity sample points; Black patch: predictions; Vertical solid line: nominal value

In contrast, when the 3 high-fidelity data points provided by the full finite element model are used for predicting by the single-level Gaussian process regression model, there is no bias in the distribution of predictions, which can been seen in Fig. 4-4(c) and 6(c). It is clear that, if more high-fidelity data points are available to the single-level Gaussian process regression model, the accuracy of its predictions may be improved to the degree that the matches become as nearly perfect as those in the examples displayed in Fig. 4-4(a) and 6(a). In reality, however, there is often no such a luxury of possession of high-fidelity data points sufficient for use. Indeed, it is very probably that a very few high-fidelity data points are acquired initially only for the purpose of verifying/validating a simulation model, but not predicting by an emulation model. When it is found that, in comparing distributions plotted in Fig. 4-4(b) and 4(c), or Fig. 4-6(b) and 6(c), detectable

discrepancies exist between predictions obtained using many low-fidelity data and predictions obtained using a few high-fidelity data, it is tempting to extract and exploit information contained in both the two data sets for better predictions.

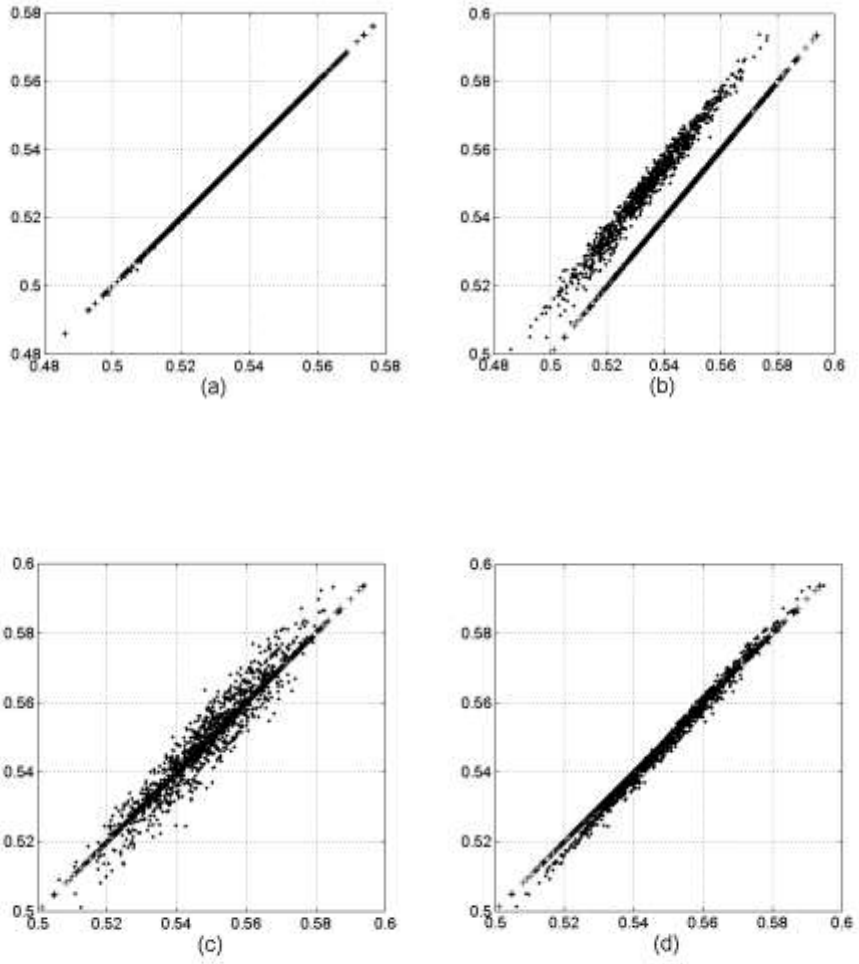

Fig. 4-6 (a) - Predictions using low-fidelity data points against low-fidelity sample points; (b) - predictions using low-fidelity data points against high-fidelity sample points; (c) - predictions using 21 low-fidelity data points against high-fidelity sample points; (d) predictions using 3 high-fidelity data points against high-fidelity sample points. Vertical axis: High-fidelity von Mises strain; Horizontal axis: low-fidelity von Mises strain. Cross: high-fidelity sample point; Dot: prediction

Our goal is to make a full use of both the 21 low-fidelity data and the 3 high-fidelity data simultaneously in a two-level Gaussian process regression model. Our expectation is such predictions should be more accurate than predictions obtained by a single-level Gaussian process regression model using only the 3 high-fidelity data. In Fig. 4-4(d) and 6(d), the predictions by the two-level Gaussian process regression model, in which the data sets at both two fidelity levels are used, are displayed the same way as in Fig. 4-4(c) and 6(c) where the predictions by the single-level Gaussian process regression model that uses only data at high fidelity level. By simple visual comparison of Fig. 4-4(c) and 4(d), or Fig. 4-6(c) and 6(d), we can see in the latter the predictions, due to more information extracted from the 21 low-fidelity data, are more concentrated around distribution of the 'exact' values which are represented by the 1000 high-fidelity sample points that, as described previously, are obtained by brute-force Monte Carlo simulation suing the full finite element model. Such a successful improvement of accuracy in predictions demonstrates the inference capabilities of Gaussian process regression can be reinforced by exploiting data at multiple fidelity levels while the concern about errors in solutions of approximation methods, such as CMS, that involves a systematic bias can also be corrected to a significant degree.

4.3.3 Estimation under uncertainty

In the above numerical examples, it is assumed that we have the convenience of possession of 'exact' values for assessing inference results, i.e., 1000 high-fidelity sample points obtained by Monte Carlo simulation at both the low- and high-fidelity level to

examine predictions by single- or two-level Gaussian process regression. As we know, this is not the case in many practical situations where high-fidelity data are either scarce or expensive. The typical situation is that, when information is obtained from a set of data at some fidelity level, only a few data at higher fidelity can be acquired for extracting additional information to appraise accuracy and reliability of it. Information In its most general sense is a set of messages that are extracted or interpreted for a specific purpose by processing crude data. As for uncertainty analysis, such information is usually expressed in the form of expected values and its neighborhoods of uncertainty, or, more formally, means and variances, which are the two kinds of statistical measures of most concern for characterizing a specific quantity of interest.

A typical realistic situation is that the exact value of a quantity of interest can never be known, and the only way to learn about it is to make inference from observed values that it takes on in physical experiments or numerical simulations. Take the sixth eigen-frequency of the exemplary structure as quantity of interest and suppose no nominal or 'exact' value but two sets of data are available, i.e. the 21 low-fidelity data and the 3 high-fidelity data. Now, the task is to estimate the expected value of the sixth eigen-frequency. Fig 5(a) displays the histogram of the predictions by the single-level Gaussian process model using the 21 low-fidelity data points only. Based on the values both of the predictions at 889 target points and the 21 data points, the expected value of the sixth eigen-frequency is calculated to be 103.67Hz. However, when comparing with data of higher fidelity, e.g., the 3 data points 92.53Hz, 96.27Hz and 98.97 Hz given by the full finite element model, we immediately realize that the expected value 103.67Hz that is

inferred by using low-fidelity data alone very probably may be far away from the nominal value 96.33Hz, which, as in most real cases, is supposed to be unknown in this example, because the predicted distribution, shown as the histogram in Fig 5(a), cannot even cover the values of the given 3 high-fidelity data. To make a more accurate inference about the value of the sixth eigen-frequency, we have two options: to use the 3 high-fidelity data alone or to use simultaneously both the 3 high-fidelity and the 21 low-fidelity data. Since the numerical cost of the two-level Gaussian processes often can be negligible compared to that of one single run of the full finite element model, it is really worthy to take the latter option to do inference if we take into consideration the improvement in accuracy that may be gained as illustrated in Section 4.2. In Fig 5(b) is shown the histogram of the predictions by two-level Gaussian processes suing both the 21 low-fidelity and the 3 high-fidelity data, which gives the expected value 96.0836Hz that is very close to the 'exact' expected value 96.0846Hz, which is obtained from the 1000 high-fidelity sample points but, again, is supposed to be unknown in this example as in most real cases.

This example demonstrates that, given a well-defined input parameter configuration space, a relatively large amount of low-fidelity data, and a relatively small amount of high-fidelity data, a quantity of interest that is subject to uncertainty, e.g., a specific eigen-frequency varying due to variations in material properties of a structure, can be estimated with guaranteed accuracy by two-level Gaussian processes. In Section 4.2, the emphasis is placed on estimations, or predictions, at individual target points and, in this sub-section, a sample constituted of these individual predictions has been further

processed, particularly in a realistic manner, to provide statistical measures for a quantity of interest, which is exactly the original motivation for Monte Carlo simulation.

4.4 Concluding remarks

Two-level Gaussian processes are introduced for simultaneously exploiting both low-fidelity data, typically in relatively large numbers, and high-fidelity data, typically in relatively small numbers, to emulate Monte Carlo simulation for uncertainty analysis in structural dynamics. CMS offer an economical approach to modeling a large complex structure while sacrificing accuracy in the data of low fidelity it generates. Single-level Gaussian process regression can predict with great efficiency but inherits all errors carried in the low-fidelity data points. Numerical examples give encouraging results that indicate two-level Gaussian processes, by using data at different fidelity levels, may not only produce efficient predictions with more accuracy at target input points, but also, as emulating Monte Carlo simulation, make better estimations of statistical measures for a quantity of interest under uncertainty. It is worthy to stress that all successful inferences are started with appropriate assumptions about underlying structures of observed data. Care should be taken to pre-examine properties of distributions of given data sets before Gaussian processes are adopted in practice. Another critical key to successful application of two-level Gaussian processes is the existence of meaningful correlation between the solutions of the two models at different fidelity levels, e.g., a full finite element model and a finite-element-based CMS model.

Chapter 5: Conclusions

Two fundamental aspects of efficient characterization of structural dynamic responses have been discussed in this thesis. Conventional approaches are based on first principles that mainly include physical mechanisms and mathematical manipulations. Order reduction techniques are a typical example of this kind of approaches. From the description of free interface CMS modeling and the numerical examples presented in this thesis, it is demonstrated that order reduction techniques can be useful in cases where physical simplifications and mathematical approximations bring forth no significant errors in results of analyses. Modern data-driven approaches take advantage of both the computer power and the well-established library of statistical methods and data analysis tools. Based on educated guessed about inner structures of datasets, this kind of methods can be very efficient for predicting target outputs of a system given sufficient data points. Gaussian process regression introduced in this thesis has shown great flexibility in data modeling.

In order to simultaneously take advantage of order reduction modeling and data modeling, two-level Gaussian process regression has been introduced in this thesis. A critical part of the multi-level data modeling is to predetermine the inner structures of the discrepancies between datasets at different fidelity levels. Gaussian distribution is not only a convenience choice but, more importantly, is the best representative distribution in many real cases where multiple sources of uncertainty exist and the application of the central limit theorem (CLT) can be well justified.

The essential processes of both the first-principle-based approaches and the data-based approaches are the same: start from prior knowledge and proceed with new observations. However, there is an apparent difference between these two categories of approaches. The first-principle-based approaches believe more in essences, if any, such as physics laws and mathematical principles, and are inclined to fit new observations into an existing subject according to its kind, but the data-based approaches are more open to observations and satisfied with phenomenological descriptions. Statistical methods offer a set of powerful data analysis tools when combined with modern computer capacities. It has become a trend that the combination of these two categories of approaches is applied to more and more engineering fields. Theoretical analysis and numerical examples in this thesis shows it also offers a promising approach to structural analysis with uncertainties.

References

Adhikari, S. and Sarkar, A., 2009, Uncertainty in structural dynamics: experimental validation of a Wishart random matrix model, Journal of Sound and Vibration, V323, pp. 802-825

Allen, M. S., Mayes, R. L., and Bergman, E. J., 2010, Experimental modal substructuring to couple and uncouple substructures with flexible fixtures and multi-point connections, Journal of Sound and Vibration, V329, pp.4891-4906

Alvarez, M. A., and Lawrence, N. D., 2011, Computationally efficient convolved multiple output Gaussian processes, Journal of Machine Learning Research, V12, pp.1459-1500

Andrieu, C., de Freitas, N., Doucet., A., and Jordan, M. I., 2003, An Introduction to MCMC for machine learning, Machine Learning, V50, pp. 5-43

Azman, K., and Kocijan, J., 2009, Fixed-structure Gaussian process model, International Journal of Systems Science, V40, pp.1253-1262

Bah, M. T., Nair, P. B., Bhaskar, A. and Keane, A. J., 2003, Stochastic component mode synthesis, 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Norfolk, Virginia

Bastos, L. S., and O'Hagan, A., 2009, Diagnostics for Gaussian process emulators, Technometrics, V51, pp.425-438

Bladh, R., Castanier, M. P., and Pierre, C., 2001, Component-mode-based reduced order modeling techniques for mistuned bladed disks-part I: theoretical models, ASME Journal of Engineering for Gas Turbines and Power, V123, pp.89-99

Bladh, R., Pierre, C., and Castanier, M. P., 2002, Dynamic response predictions for a mistuned industrial turbomachinery rotor using reduced-order modeling, ASME Journal of Engineering for Gas Turbines and Power, V124, pp.311-324

Butland, A. and Avitabile, P., 2010, A reduced order, test verified component mode synthesis approach for system modeling applications, Mechanical Systems and Signal Processing, V24, pp. 904-921

Christensen, E. R., 2005, Structural dynamic analysis of cyclic symmetric structures, 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics & Material Conference, Austin, Texas.

Craig, R. and Chang, C. J., 1977, Substructure coupling for dynamic analysis and testing, CR-2781, NASA

Craig, R. and Kurdila A., 2006, Fundamentals of structural dynamics second edition, John Wiley and Sons, Ltd

Craig, R., 2000, Coupling of substructure for dynamic analyses: an overview, 41st AIAA/ASME/ASCE/AHSIASC structures, structural dynamics, and materials conference and exhibit, Atlanta, Georgia, pp.3-14

De Gersem, H., Moens, D., Desmet, W., and Vandepitte, D., 2007, Interval and fuzzy dynamic analysis of finite element models with superelements, Computer and Structures, V85, pp. 304-319

de Klerk, D., Rixen, D. J., and Voormeeren, S. N., 2008, General framework for dynamic substructuring: history, review, and classification of techniques, AIAA Journal, V46, No. 5, pp. 1169-1181

DiazDelaO, F. A., and Adhikari, S., 2010, Structural dynamic analysis using Gaussian process emulators, Engineering Computations, V27, pp.580-605

DiazDelaO, F. A., and Adhikari, S., 2011, Gaussian process emulators for the stochastic finite element method, International Journal for Numerical Methods in Engineering, V87, pp.521-540

Dohnal, F., Mace, B. R., and Ferguson, N. S., 2009, Joint uncertainty propagation in linear structural dynamics using stochastic reduced basis methods, AIAA Journal, V47, pp.961-969

Feiner, D. M., and Griffin, J. H., 2002, A fundamental model of mistuning for a single family of modes, ASME Journal of Turbomachinery, V124, pp.597-605

Ghanem R. G., and Spanos P. D., 1991, Stochastic finite elements: a spectral approach, Springer-Verlag, New York

Gregorcic, G., and Lightbody, G., 2009, Gaussian process approach for modeling of nonlinear systems, Engineering Applications of Artificial Intelligence, V22, pp.522-533

Hasen, J., Murray-Smith, R., Johansen, T. A., 2005, Nonparametric identification of linearizations and uncertainty using Gaussian process models - application to robust wheel slip control, Joint 44th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC'05), Seville, Spain

Hinke L., Dohnala, F., Mace, B. R., Waters, T. P., and Ferguson, N. S., 2009, Component mode synthesis as a framework for uncertainty analysis, Journal of Sound and Vibration, V324, pp.161-178

Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A., 2006, Sequential Kriging optimization using multiple-fidelity evaluations, Structural and Multidisciplinary Optimization, V32, pp. 369 -382

Jaynes, E. T., 2003, Probability theory: the logic of science, Cambridge University Press, UK, Chapter 12

Jiang, X. and Mahadevan, S., 2009, Bayesian structural equation modeling method for hierarchical model validation, Reliability Engineering and System Safety, V94, PP. 796-809

Joshi, S. S. and Neat G. W., 1998, Lessons learned from multiple fidelity modeling of ground interferometer testbeds, Proceedings - SPIE The International Society for Optical Engineering, V3350, No. 1, pp. 128-138

Kang, J. H. and Kim, Y. Y., 2001, Field-consistent higher-order free-interface component mode synthesis, International Journal for numerical methods in engineering, Vol.50, pp.595-610

Kaveh, A. and Fazli, H., 2011, Approximate eigensolution of locally modified regular structures using a substructuring technique, Computer and Structures, V89, pp. 529-537

Kennedy, M. C. and O'Hagan, A., 2000a, Predicting the output from a complex computer code when fast approximations are available, Biometrika, V87, pp. 1-13

Kennedy, M. C. and O'Hagan, A., 2000b, Supplementary details on Bayesian calibration of computer codes, University of Sheffield, Sheffield, U.K. (available online)

Kennedy, M. C.and O'Hagan, A., 2001, Bayesian calibration of computer models, Journal of the Royal Statistical Society B, V63, p.425-464

Kenyon, J. A., and Griffin, J. H., 2003, Forced response of turbine engine bladed disks and sensitivity to harmonic mistuning, ASME Journal of Engineering for Gas Turbines and Power, V125, pp.113-120

Lalanne, C., 2009, Mechanical Vibration and Shock Analysis, Volume 3: Random Vibration, Second Edition, ISTE Ltd and John Wiley & Sons, Inc

Lee S.-Y. and Castanier M. P., 2006, Component-Mode-Based Monte Carlo simulation for efficient probabilistic vibration analysis, 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, Rhode Island

Lindley, D. V. and Smith, A. F., 1972, Bayes Estimates for the Linear Model, J. R. Statist. Soc., B. 34. 1., pp. 1-41

Loredo, T. J., 1990, From Laplace to Supernova SN 1987A: Bayesian Inference in Astrophysics, in Maximum Entropy and Bayesian Methods, Academic Publishers, Dordrecht, Netherlands

MacKay, D. J. C., 2003, Information theory, inference and learning algorithms, Cambridge University Press, Cambridge, UK

MacNeal, R. H., 1971, A hybrid method of component mode synthesis, Computers and structures, V1, pp. 581- 601

Mbaye, M., Soize, C., and Ousty, J.-P., 2010, A reduced-order model of detuned cyclic dynamical systems with geometric modifications using a basis of cyclic modes, ASME Journal of Engineering for Gas Turbines and Power, V132, 112502

Misawa, M., 2010, Component modal tests with additional mass and stiffness, AIAA Journal, V48, pp. 1840-1847

Mockus J., Eddy, W., and Reklaitis, G., 1997, Bayesian Heuristic Approach to Discrete and Global Optimization, Kluwer Academic Publishers, Dordrecht-London-Boston.

Mohanty, S., Das, S., Chattopadhyay, A., and Peralta, P., 2009, Gaussian process time series model for life prognosis of metallic structures, Journal of Intelligent Material Systems and Structures, V20, pp.887-96

Moyroud, F., Fransson, T., and Jacquet-Richardet, G., 2002, A comparison of two finite element reduction techniques for mistuned bladed disks, ASME Journal of Engineering for Gas Turbines and Power, V124, pp.942-952

Neal, R. M., 1998, Regression and classification using Gaussian process priors, in Bayesian Statistics 6, Oxford University Press

O'Hagan, A., 1978, Curve fitting and optimal design for prediction, Journal of the Royal Statistical Society B, V40, p. 1-42

Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., and Ferson, S., 2004, Challenge problems: uncertainty in system response given uncertain parameters, Reliability Engineering & System Safety, V85, pp. 11-19

Qian, Z., Seepersad, C. C., Joseph, C. R., Allen, J. K., Wu, C. F. J., 2006, Building surrogate models based on detailed and approximate simulations, Journal of Mechanical Design, V128, pp. 668-677

Rasmussen, C. E., and Williams, C. K. I., 2006, Gaussian processes for machine learning, MIT Press.

Rubin, S., 1975, Improved component-mode representation for structural dynamic analysis, AIAA Journal, V13, pp. 995-1006

Sivia, D. S., and Skilling, J., 2006, Data Analysis – a Bayesian Tutorial (2nd Edition), Oxford University Press

Soize, C., 2000, A nonparametric model of random uncertainties for reduced matrix models in structural dynamics, Probabilistic Engineering Mechanics, V15, pp. 277-294

Swiler, L.P., 2006, Bayesian methods in engineering design problems, Sandia National Laboratories, SAND2005 – 3249

Thacker, B. H., Doebling, S. W., Hemez, Anderson, M. C., Pepin, J. E., and Rodriguez, E. A., 2004, Concepts of model verification and validation, Los Alamos National Laboratory, USA, LA-14167-MS

Tournour M. A., Atalla, N., Chiello, O., and Sgard, F., 2001, Validation, performance, convergence and application of free interface component mode synthesis, Computers and Structures, V79, pp.1861-1876

Tran, Q. H., Quisse, M., and Bouhaddi, N., 2010, A robust component mode synthesis method for stochastic damped vibroacoustics, Mechanical Systems and Signal Processing, V24, pp. 164-181

Tustin, W., 1967, Vibration and shock tests do not duplicate service environment, Test Engineering, V18, No. 2, pp. 18-21

Whitehead, D. S., 1998, Maximum factor by which forced vibration of blades can increase due to mistuning, ASME Journal of Engineering for Gas Turbines and Power, V120, pp.115-119

Xia, Z. and Tang, J., 2010, Efficient order reduction in vibration analysis of periodic structures with uncertainties, Proceedings of SPIE, V7649

Xia, Z. and Tang, J., 2011, Characterization of structural dynamics with uncertainty by using Gaussian processes, Proceedings of ASME IDETC, DETC2011-48804